

SNO+ Waterphase Burst Principal Component Analysis

by

Philip Rost

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Physics

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

© Philip Rost, 2019

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	SNO+ Waterphase Burst Principal Component Analysis	
Name of Candidate Nom du candidat	Rost, Philip Michael	
Degree Diplôme	Master of Science	
Department/Program Département/Programme	Physics	Date of Defence Date de la soutenance October 30, 2019

APPROVED/APPROUVÉ

Thesis Examiners/Examineurs de thèse:

Dr. Clarence Virtue
(Co-Supervisor/Co-directeur de thèse)

Dr. Christine Kraus
(Co-supervisor/Co-directrice de thèse)

Dr. Jacques Farine
(Committee member/Membre du comité)

Dr. Richard Gran
(External Examiner/Examineur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

I, **Philip Michael Rost**, hereby grant to Laurentian University and/or its agents the non-exclusive license to archive and make accessible my thesis, dissertation, or project report in whole or in part in all forms of media, now or for the duration of my copyright ownership. I retain all other ownership rights to the copyright of the thesis, dissertation or project report. I also reserve the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that this copy is being made available in this form by the authority of the copyright owner solely for the purpose of private study and research and may not be copied or reproduced except as permitted by the copyright laws without written authority from the copyright owner.

Abstract

SNO+ is a kilo-tonne scale neutrino detector utilizing much of the same hardware that was used during the SNO experiment. The SNO+ experiment will be conducted using three different target media in three phases: water phase, pure scintillator phase and tellurium loaded scintillator phase. Through all phases SNO+ will be sensitive to a large neutrino burst from a nearby supernova.

Data bursts can be caused by a supernova neutrino burst or other physical phenomena such as electronic pickup, static discharge, equipment malfunctions and unintended light injection. This thesis examines data bursts which occurred during the light water phase commissioning of detector operation using a principal component analysis. The principal component analysis showed 3 major groupings within the analysed bursts: bursts during period of time with detector running issues, burst generated due to electronics break downs or light injection and bursts occurring during periods of time where the detector is operated in an abnormal running mode.

The analysis in this thesis also shows that many of the data bursts were caused by detector running issues after some initial burst event. Since the state of the detector has been improved, a repetition of this study is recommended with more recent data.

Acknowledgements

I would like to take this opportunity to thank my co-supervisors Dr. Clarence Virtue and Dr. Christine Kraus for providing the opportunity and support I needed to progress through the program. They both created an environment where I could develop both academically and professionally while serving as my mentors.

I would also like to thank specific members of the SNO+ collaboration both past and present and in no particular order: Teal Pershing, Dr. Erica Price Caden, Matt Depatie, Ian Lam, Zachariah Barnard, Pooja Woosaree, Anthony LaTorre and Eric Marzec. I think these individuals helped me with the brunt of the issues I encountered along the way and helped me sort out any silly questions that I had.

I cannot thank enough my partner Chloe Gagnon - who I promise I will marry - for her love, support and guidance during this. Without her I would not have been able to succeed. I would also thank my parents Otto and Theresa Rost for their love and support through my entire past and extending into the future.

Contents

1	Introduction	1
1.1	The Standard Model	2
1.1.1	Fundamental Particles	2
1.1.2	Fundamental Forces	4
1.2	Nuclear Decay	5
1.3	Neutrino Astrophysics	7
1.4	The SNO Project	9
2	Stars and Supernovae	13
2.1	Solar Model	13
2.1.1	pp Chain	13
2.1.2	CNO Cycle	16
2.2	Stellar Evolution	18
2.2.1	Main Sequence	19
2.2.2	Post Main Sequence	20
2.2.3	High Mass Stars Post He Burning	22
2.3	Supernovae	24
2.4	Supernova Observation with Neutrinos	26
2.4.1	Supernova Neutrino Production	27
2.4.2	Supernova Neutrino Detection	30
3	Introduction to SNO+	32
3.1	Apparatus	32
3.1.1	Liquid Scintillator	35
3.1.2	Detector Electronics	37
3.2	Experimental Goals	40
3.2.1	Neutrinoless Double Beta Decay	41
3.2.2	Solar Neutrinos	47
3.2.3	Reactor Neutrinos	48
3.2.4	Geo Neutrinos	48
3.2.5	Supernova Neutrinos	49
4	Detector Response	50
4.1	Analog Signal Chain	54
4.2	Digital Signal Chain	58
4.3	Detector Control and Event Builder	60
4.4	Burst File Production	62
5	Burstfile Analysis	67
5.1	Data Cleaning Flags	68
5.2	Data Handling	73
5.3	Representation of Burst Data in Principal Components	76

6	Results and Interpretation	81
6.1	Data Flag Distributions	81
6.2	Principal Component Analysis	94
6.3	Bursts With Significant Content	104
6.4	Removing Significant Backgrounds	106
6.5	Comparison to SNO Data	109
7	Conclusion	116
A	SNO+ Data Cleaning Flag Descriptions	125
A.1	Zero-Zero	125
A.2	Crate Isotropy	125
A.3	Fitterless Time Spread	126
A.4	Flasher Geometry Cut	127
A.5	In Time Channel Time Spread	128
A.6	Junk	128
A.7	Neck	128
A.8	OWL	129
A.9	Charge Cluster	129
A.10	QvNhit	129
A.11	QvT	130
A.12	Ring of Fire	130
A.13	CAEN	131
A.14	Muon Followers	133
A.15	Missed Muon Follower	133
A.16	Polling Cut	134
A.17	Ped Cut	134
B	Burst File Lists	136
B.1	Names of bursts in regions 1, 2 and 3	136
B.2	List of burst causes for regions 1, 2 and 3	153
C	Additional PCA Plots	163
C.1	Entire Data set PCA plots	163
C.2	Significant Bursts PCA plots	193
C.3	First Bursts PCA plots	208
C.4	All Bursts Projected Onto First Bursts PCA plots	212
D	Normalized Data Cleaning Flag Distribution Histograms	216
E	Data Analysis Scripts	225
E.1	Batch-wise Data Processing	225
E.2	ROOT Data Cleaning Flag Extraction Macro	228
E.3	Data Cleaning Flag Vector Assembly	235

E.4	Vector Normalization Script	238
E.5	Python Script to Create Desired Vector Components	240
E.6	Principal Component Analysis Script	241

List of Figures

1	Table of elementary particles.	2
2	Quark structure of a neutron.	3
3	β -decay Feynman diagram	5
4	CC interactions for beta decay	8
5	NC elastic scattering	9
6	SNO Detector Construction	12
7	The pp chain	15
8	The CNO cycle	17
9	Hertzprung-Russell diagram	19
10	Stellar hydrogen burning	20
11	Thoroughly exhausted massive star cutaway	22
12	Atomic binding energy per nucleon as a function of atomic number	23
13	SN 1994D	24
14	Supernova taxonomy decision tree	26
15	Illustration of relative neutrino penetrating power	27
16	Buras model of supernova neutrino luminosity vs time	29
17	Rendering of the SNO+ detector	33
18	Rendering of hold down rope net	34
19	Overview of the scintillator purification process	37
20	SNO+ R1408 PMT and housing	38
21	SNO/SNO+ trigger system diagram	39
22	$2\nu\beta\beta$ and $0\nu\beta\beta$ Feynman diagrams	43
23	Relative neutrino mass hierarchy	44
24	Effective neutrino mass as a function of the light neutrino mass	46
25	SNO+ Backgrounds for $0\nu\beta\beta$	47
26	XsnoED Event display	51
27	XsnoED showing a "flasher" event	52
28	SLOW Controls	53
29	Physical layout of SNO+ detector	55
30	Hamamatsu R1408 PMT schematic	56
31	PMT Hex cell assembly diagram	57
32	Master Trigger Card Analog (MTCA) block diagram	58
33	Master Trigger Card Digital (MTCD) block diagram	59
34	L2 trigger block diagram	65
35	CAEN Cut constants and bounds	73
36	FTScut distribution	82
37	Flashergeocut distribution	83
38	ITCtimespreadcut distribution	85
39	Missingcaendata distribution	86
40	Qcluster distribution	88
41	QvNhit distribution	89

42	QvT distribution	90
43	Ringoffire distribution	91
44	CAENcut distribution	92
45	Number of events per burst file distribution	93
46	Principal component 0 vs 1, representative burst sample	98
47	Principal component 0 vs 1 analysis	99
48	PCA plot with "unique" bursts highlighted	103
49	"significant" burst PCA plot	106
50	First Burst PCA plot	109
51	Projection of all bursts onto First Burst principal components	110
52	Projection of SNO+ burst data onto SNO principal components	112
53	Projection of SNO+ burst data onto SNO principal components - expanded view	113
54	SNO Principal component analysis	115
55	PCA plot of 5180 bursts, component 1 vs 0	163
56	PCA plot of 5180 bursts, component 2 vs 0	164
57	PCA plot of 5180 bursts, component 3 vs 0	165
58	PCA plot of 5180 bursts, component 4 vs 0	166
59	PCA plot of 5180 bursts, component 5 vs 0	167
60	PCA plot of 5180 bursts, component 6 vs 0	168
61	PCA plot of 5180 bursts, component 7 vs 0	169
62	PCA plot of 5180 bursts, component 2 vs 1	170
63	PCA plot of 5180 bursts, component 3 vs 1	171
64	PCA plot of 5180 bursts, component 4 vs 1	172
65	PCA plot of 5180 bursts, component 5 vs 1	173
66	PCA plot of 5180 bursts, component 6 vs 1	174
67	PCA plot of 5180 bursts, component 7 vs 1	175
68	PCA plot of 5180 bursts, component 3 vs 2	176
69	PCA plot of 5180 bursts, component 4 vs 2	177
70	PCA plot of 5180 bursts, component 5 vs 2	178
71	PCA plot of 5180 bursts, component 6 vs 2	179
72	PCA plot of 5180 bursts, component 7 vs 2	180
73	PCA plot of 5180 bursts, component 4 vs 3	181
74	PCA plot of 5180 bursts, component 5 vs 3	182
75	PCA plot of 5180 bursts, component 6 vs 3	183
76	PCA plot of 5180 bursts, component 7 vs 3	184
77	PCA plot of 5180 bursts, component 5 vs 4	185
78	PCA plot of 5180 bursts, component 6 vs 4	186
79	PCA plot of 5180 bursts, component 7 vs 4	187
80	PCA plot of 5180 bursts, component 6 vs 5	188
81	PCA plot of 5180 bursts, component 7 vs 5	189
82	PCA plot of 5180 bursts, component 7 vs 6	190
83	FTScut flags/event distribution	216

84	Flashergeocut flags/event distribution	217
85	ITCtimespreadcut flags/event distribution	218
86	Missingcaendata flags/event distribution	219
87	Qcluster flags/event distribution	220
88	QvNhit flags/event distribution	221
89	QvT flags/event distribution	222
90	Ringoffire flags/event distribution	223
91	CAENcut flags/event distribution	224

List of Tables

1	Neutrino reactions that can be detected inside of D_2O and H_2O	11
2	Reactions creating neutrinos in the p-p chain	15
3	Table of selected double β -decay candidate isotopes	42
4	Representative sample principal component vectors	95
5	Variance of each calculated representative sample principal component	96
6	Relative weight of data cleaning flags for each vector component of principal component 0, 1 and 1-0	97
7	Probable cause of bursts during runs producing bursts in concentrated groupings. Continued in Table 8.	101
8	Probable cause of bursts during runs producing bursts in concentrated groupings. Continued from Table 7.	102
9	Percentage of bursts remaining, requiring a cleaning cut pass rate chosen from the representative sample of every 10^{th} burst file.	105
10	Original principal components vs significant physics data principal components	105
11	Original principal components vs "First Burst" principal components	108
12	Burst data principal components during SNO	111
13	Modification of SNO principal components to suit SNO+ burst data	111

1 Introduction

The field of Physics came to exist more than 2000 years ago and has developed into a method of deducing reasonable processes in the form of mathematical laws to explain observable phenomena such as the motion of the sun, moon and stars. This was a new approach since the prevailing school of thought was that the changes in the cosmos were due to supernatural or spiritual mechanisms. This desire to explain astronomical, terrestrial and microscopic phenomena has made way for significant advancement in Mechanics, Thermodynamics, Electromagnetic Theory, Cosmology and others with the principles of each overwhelmingly spilling over into the studies of the other.

As it happens, the more we have learned about the mysteries of the universe the more questions we find to ask. These questions seem to grow more complex with each discovery that is made. Larger and more precise apparatuses implementing novel mechanisms have become necessary to answer these questions. This progression has led to an unlikely method of probing one of the most energetic phenomena known to man – that is detection of supernova neutrinos in order to study the supernova mechanism itself.

This idea that the smallest constituent of matter can reveal secrets of one of the largest and most energetic phenomena observed by man falls into the branch of Particle Astrophysics. Particle astrophysicists use their knowledge and understanding of fundamental particles and forces in order to explain phenomena at a universal scale.

into complex systems. The 12 fermions are subdivided into two categories: quarks and leptons.

The quarks are bound together into groups of two or three called mesons or baryons respectively. Neutrons and protons are examples of baryons and are comprised of some combination of 3 up and down quarks. Figure 2 shows an illustration of quarks bound by gluons to form a neutron. More exotic baryons can also be created with other combinations of the six quarks. Mesons are comprised of quark – antiquark pairs and are generally (with some exceptions) less stable than baryons [2].

The leptons consist of six neutrinos: particle and anti-particle variants of electron, mu and tau flavours and the particle and anti-particle variants of the electron, muon and tau leptons.

Elementary bosons are the propagators of interactions between particles. The photon is the propagator of the electromagnetic force, the gluon is the propagator of the strong force while the W^\pm and Z^0 bosons are the propagators of the weak force. The Higgs boson is the newest addition to the Standard Model after being confirmed by experiments at CERN in 2012.[4]

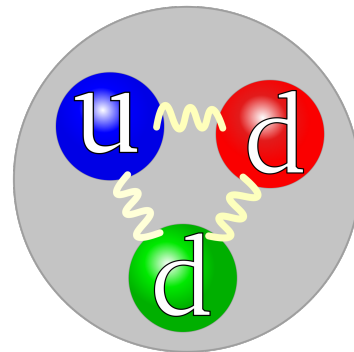


Figure 2: Neutrons are composed of two down quarks and a single up quark as shown in the figure. The three quarks are bound together by gluons. [3]

1.1.2 Fundamental Forces

Much like the set of elementary particles, their interaction mechanisms are confined to a finite set. These are commonly referred to as the four fundamental forces; the strong force, weak force, electromagnetic force and gravity. Together, these forces bind quarks together into atoms, atoms together into molecules, molecules together into visible matter and all visible matter into all of the matter in the universe.

As was mentioned in the previous section, the strong force is mediated between particles by the gluon and is responsible for binding groups of quarks together. This allows for protons and neutrons, which are the basis of the atomic nucleus, to be formed. This is the strongest of the fundamental forces acting at a relatively short range of $\sim 1\text{fm}$.[\[2\]](#)

The W^\pm and Z^0 bosons are the carriers of the weak force. The weak force is primarily involved in interactions where one fermion is changed into some configuration of other fermions. An example of this process is beta decay, where one of the down quarks of a neutron changes into an up quark. The exchange of the W^- boson results in the production of an electron and an electron anti-neutrino. The Feynman diagram for beta decay can be seen in Figure 3.

Electromagnetism is carried out between particles by photons. This is the force many of us are most familiar with as it acts on objects readily at the human scale. This force is usually observed through the interactions of charged particles exhibiting the familiar attractive or repulsive forces associated with classical studies of electromagnetic theory. It also plays a significant role in the kinematics of nuclear and atomic systems.

Gravitation is the force responsible for the long range interaction between massive bodies. At the scale of the atom however gravitation, though present, is the weakest of the fundamental forces by a factor of 10^{-39} . A mediating particle of the gravitational force has not yet been discovered though theories of quantum gravity predict that such a particle could exist and is commonly referred to as the graviton.

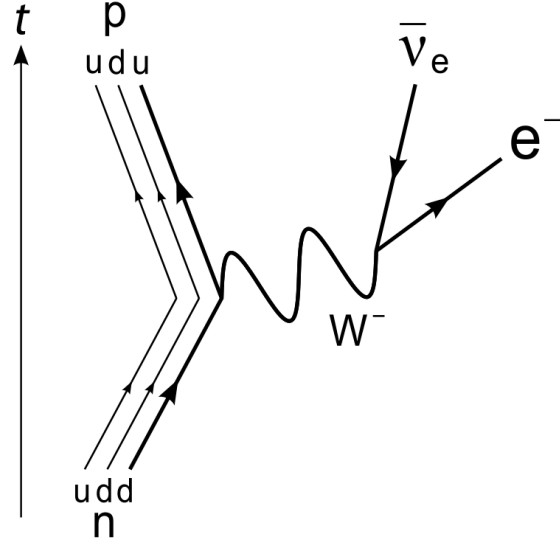


Figure 3: β -decay Feynman diagram showing the interaction of a W^- boson's role in the process. [5]

1.2 Nuclear Decay

Within the framework of the Standard Model it is possible to transform one progenitor into some combination of progeny. At the nuclear scale this can exist as a naturally occurring process which we call nuclear decay. The three most common types of nuclear decay are α -decay (emission of a ${}^4\text{He}$ nucleus) β -decay (emission an electron or positron and a neutrino) and γ -decay (emission of one or many γ or photons). Decay processes occur successively forming a decay chain until eventually stable isotopes are formed. The Table of Nuclides gives information such as stability and half-lives of the known nuclei.

β -decay (shown in Figure 3) is the most relevant decay mode to this thesis. At first sight, the emission of an electron from a nucleus which does not contain any electrons might seem difficult to understand. The interactions of neutrinos and nucleons through the weak force

however provide a path by which β -decay can be understood.

In the first half of the 20th century there was a great deal of mystery surrounding β -decay as the measured energy spectrum was continuous instead of mono-energetic. Wolfgang Pauli hypothesized that another particle lighter than the electron was playing a role in this process [6]. He proposed that the neutrino – an uncharged and "nearly" massless particle – would carry some of the energy of the transition with it. This would explain the apparent non-conservation of energy in β -decays. Because of this proposition followed by experimental confirmation, the common β -decay equations since involve the neutrino[2]:

$$\begin{aligned} &\beta^- \text{ decay} \\ n &\longrightarrow p + e^- + \bar{\nu}_e \end{aligned} \tag{1.1}$$

$$\begin{aligned} &\beta^+ \text{ decay} \\ p &\longrightarrow n + e^+ + \nu_e \end{aligned} \tag{1.2}$$

$$\begin{aligned} &\text{inverse } \beta \text{ decay} \\ \nu_e + p &\longrightarrow e^+ + n \end{aligned} \tag{1.3}$$

$$\begin{aligned} &\text{electron capture} \\ e^- + p &\longrightarrow n + \nu_e \end{aligned} \tag{1.4}$$

The previous equations for β -decay distinguish between two species of electron neutrinos the electron anti-neutrino $\bar{\nu}_e$ and the electron neutrino ν_e . We know now that there are 6 species of neutrinos in total spanning the three lepton flavors and both particles as well as

anti-particles.

The entire set of 12 leptons each have a flavour dependent lepton number associated with them. For leptons and anti-leptons, the lepton numbers are +1 or -1, respectively. We then impose that lepton numbers of dissimilarly flavoured particles must be treated separately from each other. In order to conserve lepton number then, the sum of all electron flavour, mu flavour and tau flavour lepton numbers must individually be equal to each other before and after the interaction takes place [2].

1.3 Neutrino Astrophysics

Optical astronomy has been the staple of understanding the cosmos throughout history. Over time, tools have been created that can not only magnify the visible image of stellar bodies but also probe beyond the visible light spectrum using optical, radio, and gamma telescopes. Optical astronomy has limited capabilities when determining the inner workings of a star and it can only reveal details about the processes near the surface of the star in question.

Neutrinos on the other hand have a mean free path much greater than that of photons in matter extending to distances of several light years. The star in question is thus virtually transparent to the neutrino allowing the properties of a detected neutrino to be the vessel of discovery for the inner workings of a star. Neutrinos emitted in the core of the sun can travel to earth virtually unimpeded where photons emitted in the core of a star can take tens to hundreds of thousands of years to escape.

Presently there are several different neutrino detectors which have been built to exploit the ability neutrinos have to penetrate matter and make observations of particular processes

based on the properties of the emitted neutrinos. The form which these detectors take is typically some variation of an instrumented target material which will produce a signal in the eventuality that a neutrino interaction occurs within the instrumented volume.

Since neutrinos are uncharged leptons, their only interaction mechanism is through the weak force by exchanging bosons with other particles. The W^\pm bosons are charged carriers of the weak force and so interactions where a W^\pm is exchanged are called charged current (CC) interactions. Interactions where a Z^0 boson is exchanged are called neutral current (NC) interactions.

These interactions are easiest to picture in the form of a Feynman diagram. We can see in Figures 4 and 5 that the primary difference between the CC and NC interactions is in the charge of the exchanged bosons.

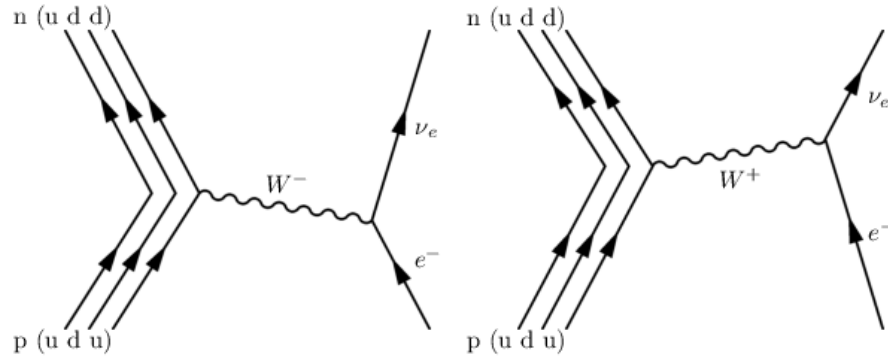


Figure 4: Feynman diagrams showing CC interactions for β -decay. These interactions involve the charged W^\pm boson[5]

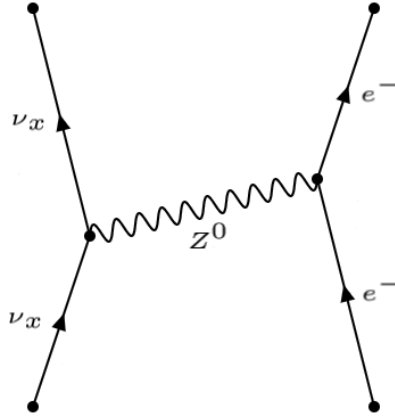


Figure 5: Feynman diagram showing the NC interaction of electron elastic scattering. This interaction involves the exchange of the uncharged Z^0 boson.

1.4 The SNO Project

The Sudbury Neutrino Observatory (SNO) was a heavy water Cherenkov neutrino detector constructed 2 kilometers below the surface of the earth at Creighton mine in Sudbury, Ontario, Canada [7]. The results of the SNO experiment contributed to solving the “solar neutrino problem” which was a significant undetected portion of the predicted solar neutrino flux. Arthur B. McDonald was a co-recipient of the 2015 Nobel Prize in Physics for the research.

SNO was designed as a large scale neutrino detector which could detect both charged current and neutral current interactions in order to probe the solar ν_e flux as well as the total solar neutrino flux. The 2 km of norite rock overburden provided shielding decreasing the background from cosmic radiation enough that the rare neutrino interactions could be detected. The ultimate implication of the measured solar neutrino flux was that the neutrino flavours could oscillate, proving that neutrinos themselves were not massless.

A neutrino can be viewed as either a linear combination of flavour or mass eigenstates that

may evolve as it propagates through space. Neutrinos are created through weak interactions as pure weak flavour eigenstates, propagate as a linear combination of mass eigenstates with quantum phases that evolve independently for each mass component, and interact through the weak interaction. A consequence of the propagation of superposed mass states is that the probability of a given neutrino interacting as a particular flavour oscillates as a function of the distance traveled. In addition to these neutrino oscillations, known as vacuum oscillations, there is a further effect that modifies neutrinos as they pass through matter - a phenomenon known as the Mikheyev–Smirnov–Wolfenstein (MSW) effect. This rich combination of neutrino oscillation phenomena has resulted in a diverse set of neutrino experiments being able to measure most of the parameters that determine the detailed behaviour of neutrinos.

SNO consisted of a 12 m diameter acrylic sphere filled with 1000 tonnes of heavy water surrounded by a 18.5 m diameter photomultiplier tube array (PSUP) which were submerged in ultra-pure water to shield the detector from backgrounds emitted in the surrounding rock [7]. Figure 6 shows the PSUP during the construction of SNO with the acrylic vessel visible in the background. This arrangement was different than other water Cherenkov detectors because the deuterium would allow for detection of NC interactions. The heavy water target was instrumental determining that neutrinos oscillate between flavours as they travel through the sun. The deuterium (^2H) provided a target capable of interacting with electron neutrino through a CC interaction as well as non electron flavoured neutrinos and anti-neutrinos through the NC interactions as can be seen in Table 1.

In addition to solving the solar neutrino problem, SNO was capable of detecting supernova neutrinos in the event that a supernova occurring within the Milky Way galaxy. The

data would be quite different in quantity from the data taken during normal running as the interaction rate would be significantly greater and depend directly on the distance at which the supernova occurred. A typical data rate for solar neutrino interactions would be on the order of a few a day where a supernova signal could cause hundreds or thousands of interactions in the span of a few seconds. The sudden burst of light in the detector volume could easily overwhelm the detector electronics as their primary design was intended for the rates that would be produced by a typical solar neutrino flux. SNO had special provisions to allow for the increased data rate that would occur in the event of a supernova neutrino burst.

Reaction	Reaction Type	D_2O	H_2O
$\bar{\nu}_e + p \longrightarrow n + e^+$	CC		•
$\nu_e + d \longrightarrow p + p + e^-$	CC	•	
$\bar{\nu}_x + d \longrightarrow \bar{\nu}_x + p + n$	NC	•	
$\nu_x + d \longrightarrow \nu_x + p + n$	NC	•	
$\bar{\nu}_x + e^- \longrightarrow \bar{\nu}_x + e^-$	Elastic Scattering	•	•
$\nu_x + e^- \longrightarrow \nu_x + e^-$	Elastic Scattering	•	•

Table 1: Neutrino reactions that can be detected inside of D_2O and H_2O

The recovery and reconstruction of supernova burst data was a priority for SNO and required stress testing of the detector to determine its supernova readiness. The stress testing was done by lowering a physical light source into the detecting volume and pulsing it quickly to simulate what might occur in the case of a supernova. This was known as the supernova calibration source and has since been re-designed for SNO's successor experiment, SNO+, with a specified dynamic range of 0-60 MeV energy deposition through pulses of length on the order of 20 ns [8].

This thesis outlines a study that was conducted using bursts of data which were recorded during and immediately after the water commissioning of the SNO+ detector. The data was generated during a period of time where the detector was filled with ultra-pure water and a similar data output to SNO would be expected. A statistical study of the bursts showed that many of them could be classified depending on their exhibited characteristics.

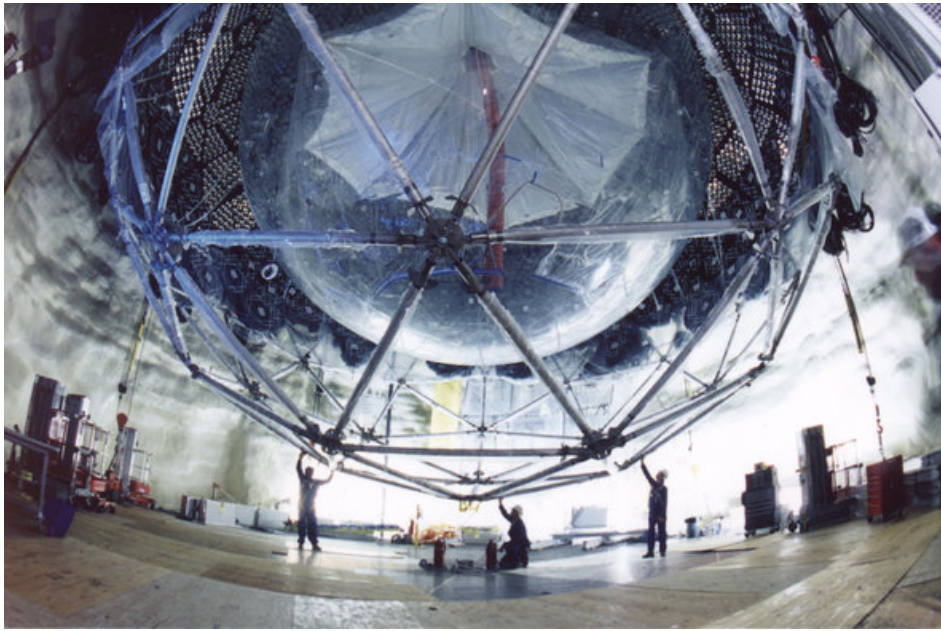


Figure 6: Pictured is the construction of the lower hemisphere of the photomultiplier tube support structure for SNO [9]

The majority of the SNO detector has been re-purposed since 2012 for a successor experiment called SNO+. SNO+ employs the cavern, acrylic vessel, photomultiplier array and much of the detector electronics with a new scintillating target to study whether or not neutrinos could be their own anti-particles. Like SNO, SNO+ will also be sensitive to a supernova neutrino signal but favoring different neutrino interactions and with a much greater light output than SNO [10]. SNO+ will be discussed in more detail in chapter 3.

2 Stars and Supernovae

Many neutrino detectors are capable of detecting neutrinos generated by nuclear processes within the sun and by more energetic events such as supernovae. The neutrino fluxes from both of these sources can vary drastically with respect to each other. The detectors are thus designed with both of these sources in mind. This section will explore some specifics of how neutrinos can be produced in both the sun and in supernovae.

2.1 Solar Model

At the center of our solar system is a massive sphere of burning gas radiating enough energy to warm the earth 150 million kilometers away through the vacuum of space. This radiant energy is produced by a variety of different nuclear reactions which ultimately fuse protons into helium nuclei. A balance of thermonuclear pressure and gravitational force is fuelled by specific cyclical chains of nuclear reactions depending on the mass of the star. For stars less than $1.3 M_{\odot}$ this equilibrium is dominated (99%) [11] by the proton proton (pp) chain. For stars greater than $1.3 M_{\odot}$ this equilibrium is dominated by the CNO cycle (carbon, nitrogen, oxygen).

2.1.1 pp Chain

The pp chain refers to one of the two mechanisms by which a star can convert hydrogen into helium. There are three main branches of the pp chain that describe most of the reactions that take place in order for this fusion to occur. They are typically labeled as pp I, pp II and pp

III with pp I accounting for more than 84% of the fused ^4He produced by the pp chain. In addition to these three main branches, there exists an additional predicted branch called the pp IV branch or the hep branch.

The pp chain begins with one of two reactions both producing a deuteron by fusion of two protons. After this, the deuteron fuses with another proton forming ^3He . At this point the reaction can branch off in three different ways (see Figure 7).



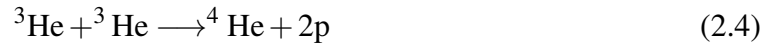
or



then ^3He is formed by:



The pp I chain will occur $\sim 85\%$ of the time where the created ^3He will fuse with another ^3He creating ^4He and releasing 2 protons:



This is the primary path of nuclear fusion within our sun and other stars that have masses less than $1.3 M_\odot$ [13]. There are other branches however that bear significance to this thesis as they result in the creation of neutrinos which can be detected by SNO+. A summary of the

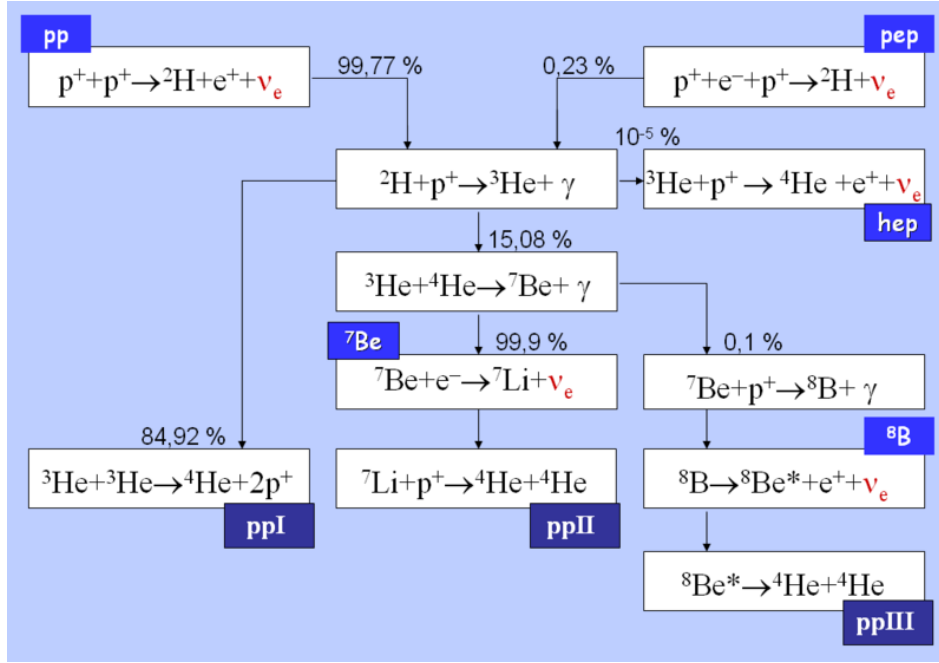


Figure 7: The pp chain is the dominant mechanism of converting hydrogen into helium in stars whose masses are less than $1.3 M_{\odot}$. Branching ratios of the various branches for our sun are labelled above each reaction.[12]

reactions from the p-p chain that produce neutrinos is shown in Table 2.

Reaction	Neutrino Energy	Neutrino Type
$p^+ + p^+ \rightarrow {}^2\text{H} + e^+ + \nu_e$	up to 0.5 MeV	pp neutrino
$p^+ + e^- + p^+ \rightarrow {}^2\text{H} + \nu_e$	1.44 MeV	pep neutrino
${}^3\text{He} + p^+ \rightarrow {}^4\text{He} + e^+ + \bar{\nu}_e$	up to 18.8 MeV	hep neutrino
${}^7\text{Be} + e^- \rightarrow {}^7\text{Li} + \nu_e$	0.861 MeV / 0.383 MeV	${}^7\text{Be}$ neutrinos
${}^8\text{B} + \rightarrow {}^8\text{Be}^* + e^+ + \nu_e$	up to 14 MeV	${}^8\text{B}$ neutrinos

Table 2: Reactions creating neutrinos in the p-p chain

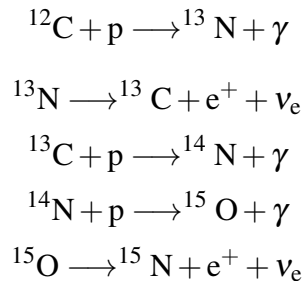
Despite the difference in contribution to the total solar neutrino flux from the main pp chain branches and the less dominant branches cannot be ignored since the produced neutrino's interacting energies coincide with the physics goals of SNO+ and so constitute a background. [14].

2.1.2 CNO Cycle

The CNO cycle (carbon - nitrogen - oxygen) is another mechanism by which stars can convert hydrogen into helium (shown in Figure 8). It is the dominant process of energy production in stars whose mass is greater than $1.3 M_{\odot}$ because the reactions require higher temperatures in order to be self-sustaining. Here carbon, nitrogen and oxygen act as catalysts undergoing successive and repetitive proton capture and subsequent nuclear decays in order to eventually convert four protons into a single helium atom emitted from the final reaction of the cycle.

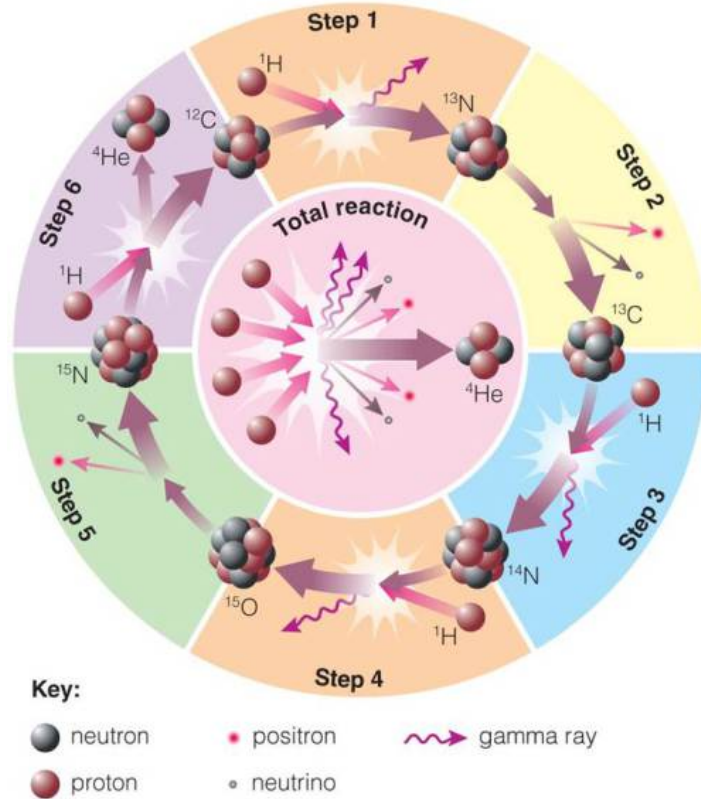
Similarly to the pp cycle many non-dominant reaction chains can exist throughout the different phases of the main CNO cycle. These chains are the CNO-I, CNO-II, CNO-III, CNO-IV and in extreme temperature the hot CNO chains HCNO-I, HCNO-II, HCNO-III. They are separated into "hot" and "cold" based on the energies required in order for the reaction to occur.

The primary CNO-I cycle begins with a proton capture on a ^{12}C nucleus. The created ^{13}N then β -decays into a ^{13}C nucleus once again undergoing a 2 proton captures to become first ^{14}N then ^{15}O . The ^{15}O will then β -decay into ^{15}N before finally undergoing α -decay to produce ^4He and ^{12}C where the cycle will begin again:





The CNO cycle and pp chain are the mechanisms through which neutrinos are produced by a star. The sun produces neutrinos primarily through the pp chain although the division between one process or the other occurring is not completely binary and a blend of the two processes will occur with one process dominating the other. The processes yield "solar neutrinos" which can be detected.



Eventually the star will use its supply of hydrogen through both of these mechanisms and the process

Figure 8: The predominant chain of the CNO cycle is outlined in the Figure. Several other chains can occur but do so far less often than the main CNO cycle and are not pictured. [15]

of stellar death will begin. For stars with mass greater than $8 M_{\odot}$ stellar death culminates in a supernova explosion.

The conversion of hydrogen into helium by the pp chain and CNO cycle give us an idea of how the fusion of heavier elements would occur during the next phases of stellar evolution. Stellar evolution and supernovae are covered in detail in section 2.2 and 2.3.

2.2 Stellar Evolution

The finite lifetime of a star depends on its mass. The processes occurring within the star change over time, moving it through the different phases of its life cycle. Generally the lifetime of lighter stars is much longer than the lifetime of more massive differing in length from tens of billions of years to just a few million, respectively. Each star maintains hydrostatic equilibrium of the inward gravitational force and the outward pressure from the fusion reaction of:



Stars can be classified into groups based on their optical luminosity and colour temperature. The Hertzsprung-Russel diagram (Figure 9) is a plot of stars where the colour temperature and luminosity have been determined. In this representation, the colour temperature is indicative of the mass of the star.

Most stars after they are born from collapsing clouds of gas fall somewhere on the distribution extending from the top left to the bottom right of the Hertzsprung-Russel diagram known as the main sequence. After the star runs out of fuel and can no longer sustain its hydrostatic equilibrium, the star will progress away from the main sequence eventually arriving at an end phase of stellar death.

Our sun falls on the main sequence, and as such is a very average star. The main sequence represents stars that fall into the longest period of stability as the life of the star progresses. During this phase, stars traverse the main sequence maintaining a dynamic equilibrium of thermonuclear pressure and gravitational binding forces.

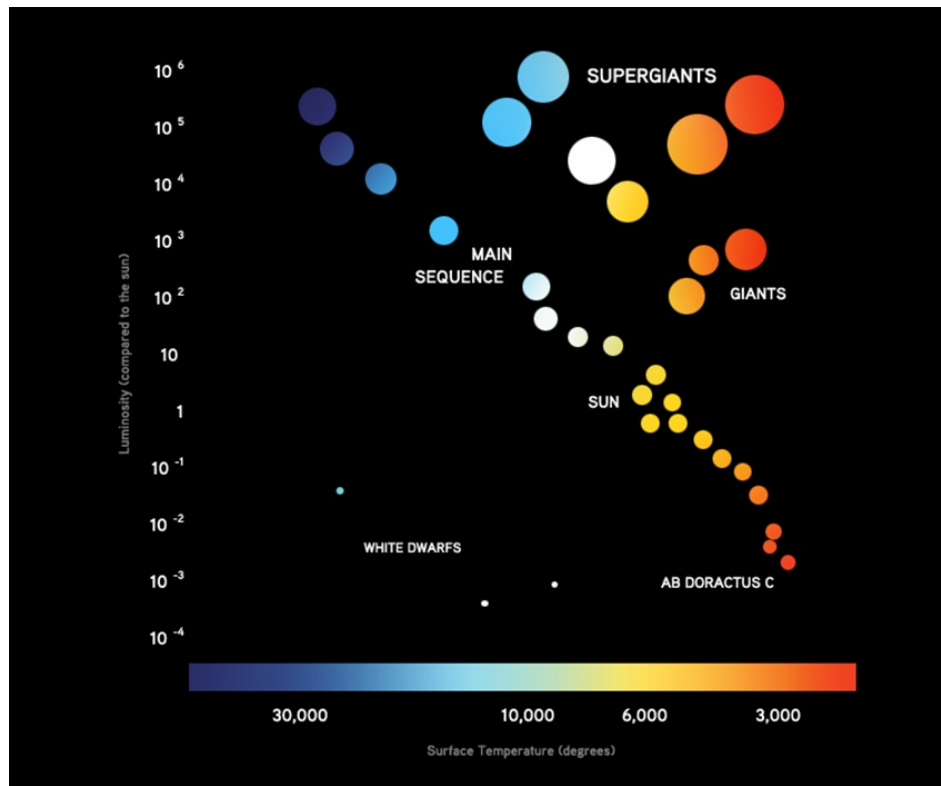


Figure 9: Hertzsprung-Russell diagram of mass (colour temperature) vs luminosity. The main sequence is where stars progress for most of their life. Once their fuel has been exhausted they move away from the main sequence in a direction depending on their mass. ESO/José Francisco (josefrancisco.org)[16]

2.2.1 Main Sequence

After a star has been formed and stabilizes it reaches the main sequence. Typically this process will take on the order of 10 million years. It is here on the main sequence that the star will spend 90% of its life. The exact position on the main sequence of the Hertzsprung-Russell diagram is determined by the mass of the star, with the most massive stars being up and to the left while the least massive are low and to the right. Since the life of the star is dependent on its mass, we know that the stars located in the top left of the Hertzsprung-Russell diagram will live the shortest lives (10 million years) while the least massive stars in the bottom right

corner will live much longer (100 billion years). After the star runs out of hydrogen to burn, it will begin to collapse under the dominant force of gravity. This will cause the star to move away from the main sequence to the beginning phases of its death.

The average life for stars in the same region of the main sequence as our own sun is about 10 billion years. Our sun is about 4 billion years old making it approximately 40% of the way through its life. After its lifetime has elapsed it too will progress away from the main sequence through to its own stellar death. [17]

2.2.2 Post Main Sequence

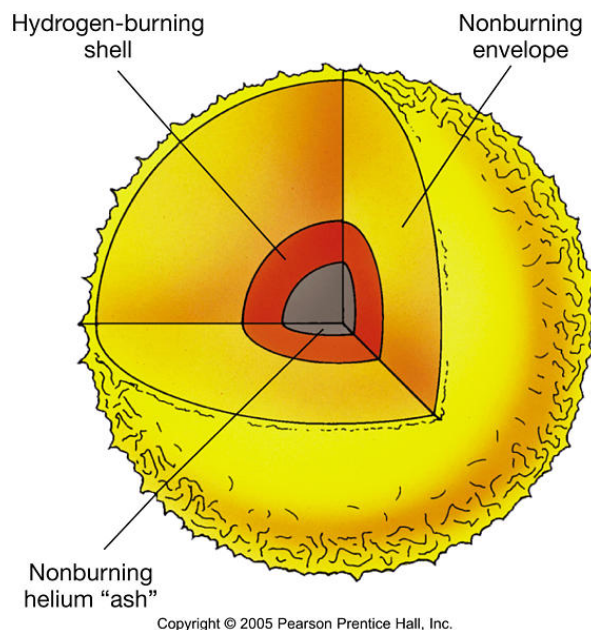


Figure 10: Stellar hydrogen burning shown as a cutaway diagram of a star. Once the helium core starts to form the star begins to expand and cool while increasing in optical luminosity. [18]

The cessation of hydrogen burning in a star allows gravitational pressure to contract the star. This contraction in turn increases the kinetic energy of the atoms composing the star and thus its temperature until hydrogen fusion can once again occur surrounding a helium core (see Figure 10). Hydrogen shell burning occurs more rapidly than hydrogen core burning, heating up the star and causing it to expand.

This expansion moves that star into the red giant area of the Hertzsprung-Russel diagram, up and to the right of where it was on

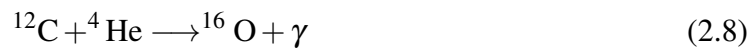
the Main Sequence. This is because the rapid burning and expansion will increase the total optical luminosity of the star, but the expansion will decrease the overall colour temperature. The star has become both redder and larger hence the name red giant.

The expansion of a star during this phase is very dramatic. It is expected that the diameter of our sun when it transitions to the red giant phase will extend to $250\times$ its current size. Despite this expansion, the size of the helium ash core does not change. The helium ash produced from the hydrogen burning shell rains down into the helium core from the hydrogen burning shell.

Eventually, the persistent accumulation of helium ash will increase the temperature and mass of the helium core until a helium flash occurs. At this point, the core has heated enough that it can fuse helium into carbon and oxygen:



and



This initiates the helium core burning of the star, which progresses in a similar fashion to the hydrogen core burning of a main sequence star. On the Hertzsprung Russel diagram, the star moves away from its red giant phase, down and to the left back towards the main sequence stars but without reaching the main sequence. Here a new hydrostatic equilibrium is reached where the gravitational pressure is opposed by the helium core and hydrogen shell burning. At this point the fate of the star depends on its mass. For the sake of this thesis

we will focus on high mass stars as they are more likely to produce supernovae. It is worth noting that in general low mass stars generally expand and cool shedding their outer layers until they become nebulae and white dwarfs. [17]

2.2.3 High Mass Stars Post He Burning

In this context, high mass stars are those which have a mass greater than $8 M_{\odot}$. For these stars, a succession of core and shell burning stages will continue until an iron ash core has been produced and fusion can no longer occur. Figure 11 shows the fully populated shell structure of a depleted high mass star.

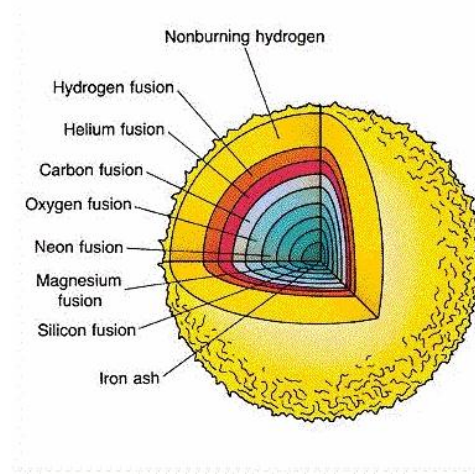


Figure 11: Thoroughly exhausted massive star cutaway. Providing the star is sufficiently massive its successive fusion of metals will cease with an iron core. [18]

This increased number of shells populates relatively quickly (hundreds of years) compared to the initial hydrogen-helium

shell due to the \sim asymptotic nature of energy released through fusion as the atomic number of the produced isotope climbs to 56. This is because the energy yield of the fusion of heavier elements and the absolute number of fusions that can occur decreases as the shells progress through heavier elements. A plot of binding energy against atomic number shown in Figure 12, clearly shows a maximum at ^{56}Fe . For a fusion reaction to be energetically favorable, the binding energy per nucleon of the product nucleus must be some amount greater

than the parent nuclei. Beyond ^{56}Fe it is no longer energetically favorable for any more fusion to occur since the binding energy of produced nuclei is less than the binding energy of ^{56}Fe .

Once the star has reached the end of its fuel supply and it can no longer support itself with the energy that is produced through fusion, the star will begin a violent struggle between electron degeneracy pressure and gravitational pressure culminating in an implosion which then results in the shedding of the stars outer layers in the form of a supernova [17]. The details of supernovae are covered in the next section of this thesis.

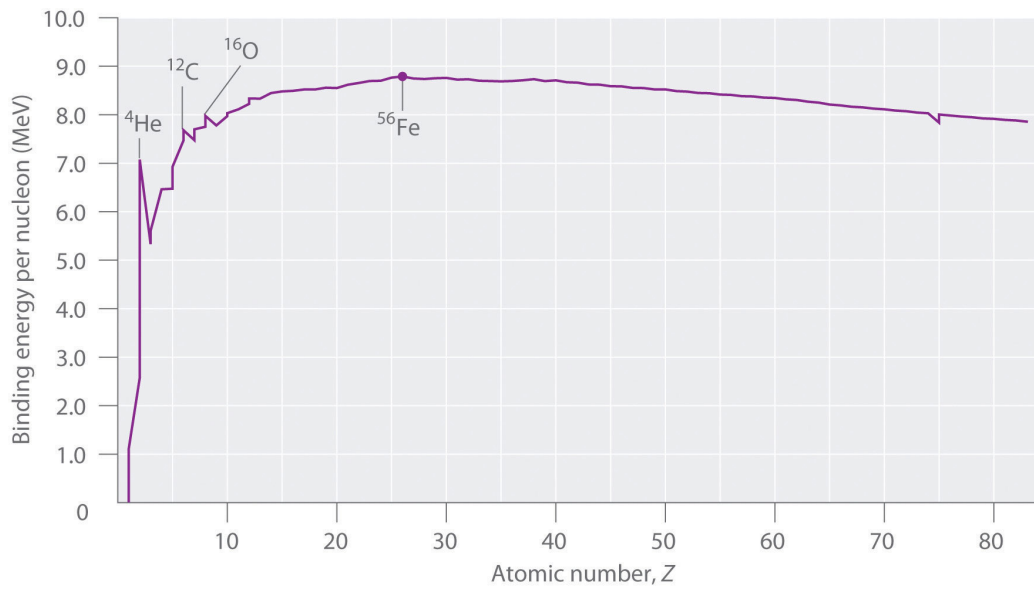


Figure 12: Binding energy of a nucleus vs atomic number. The change of slope at ^{56}Fe indicates the reason why fusion beyond the iron nucleus is not possible. The negative slope beyond ^{56}Fe indicates that all heavier nuclei that could be produced through a fusion reaction would not be energetically favorable since their binding energy per nucleon is less than that of ^{56}Fe . [19]

2.3 Supernovae

A supernova is the astronomical event marking the final stage of a star's life after it has consumed all of its fuel, succumbing to gravitational collapse. It is during this phase that the star has lost its ability to maintain a hydrostatic equilibrium between thermonuclear and gravitational pressure. The star then releases its gravitational binding energy in the form of a massive explosion (Figure



13).

Figure 13: SN 1994D (bright spot on the lower left), a Type Ia supernova outshining its home galaxy NGC 4526. [20]

Only a handful of supernovae have ever been observed and recorded within the Milky Way galaxy in the past thousand years although projections based on optical astronomy of other galaxies predict that roughly 3 galactic supernovae would be expected per century. Today, with the level of technology available to astronomers it is obviously desirable that the modern astronomer be able to observe the supernova when it happens in order to further understand the supernova mechanism itself.

The current model predicts that at the onset of a supernova (not yet observed), the optical luminosity increases to a peak over ~ 80 days before slowly decreasing over ~ 2 years. The shape of this light curve and its associated absorption spectrum serves as the basis for the

categorization of supernovae. Categorization of supernovae follows from a decision tree of observable parameters. The first classification of supernovae is given by whether or not the supernova exhibits hydrogen absorption lines in its spectra. Presence of these lines indicates a Type II supernova where a lack of these lines indicates a Type I supernova.

Type I supernovae are then further categorized based on the presence of silicon lines in their absorption spectra. Those with silicon lines are called Type Ia while those without are once again categorized by the presence of helium lines. Type Ib supernovae have helium lines but no silicon lines, Type Ic do not have silicon or helium lines. Type II supernovae are classified into Type IIP if their light curve plateaus or Type IIL if their light curve decreases linearly over time. A Type I Ib is one that shows both hydrogen and helium lines in the absorption spectrum. The Type Ia supernova is a thermal runaway supernova which of course refers to runaway nuclear fusion progressing beyond the fusion of carbon. All Type II supernovae as well as Type Ib and Type Ic supernovae are core-collapse supernovae which are expected to provide the most interesting neutrino spectra to study. This is all summarized by Figure 14 [21].

The supernova mechanism itself has proven to be an elusive entity to study since supernovae are a relatively rare occurrence. Further challenges given their distance and unpredictability require special measures be taken in order to study the supernova mechanism as it happens. Current models predict that the vast majority of the gravitational binding energy of the star is released in the form of neutrinos emitted early (first few hundred milliseconds [22]) during the progression of the supernova. These neutrinos would provide information about the supernova explosion mechanism as well as an early warning so that the early light

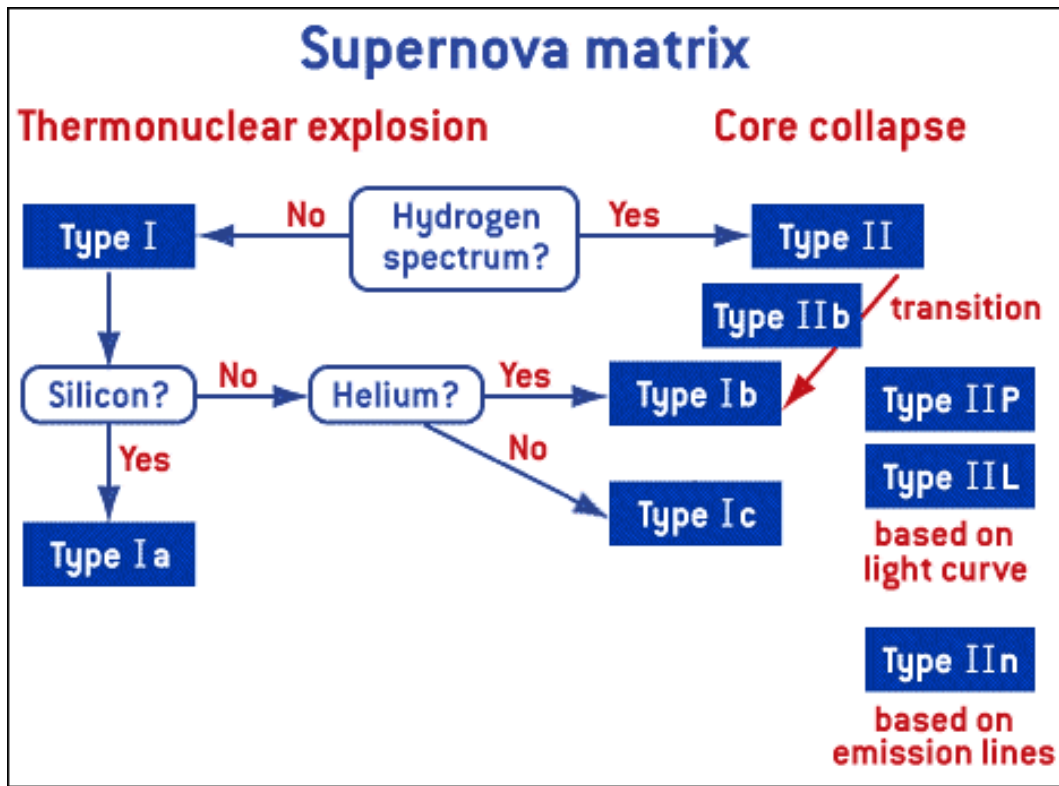


Figure 14: The above decision tree outlines the parameters used to categorize a specific supernova. Metallicity and characteristics of the light curve are used to categorize the supernova.

signal from the supernova could be observed.

2.4 Supernova Observation with Neutrinos

The neutrinos emitted by a supernovae are weakly interacting particles and can travel through matter virtually unimpeded as illustrated by Figure 15. Since neutrinos can pass through large bodies without interacting at all, they can be used to probe deep core processes in stars and supernovae.

This weak interaction however also makes detecting the particles challenging. Just as they can pass through entire planets without interacting, the neutrino may also pass through

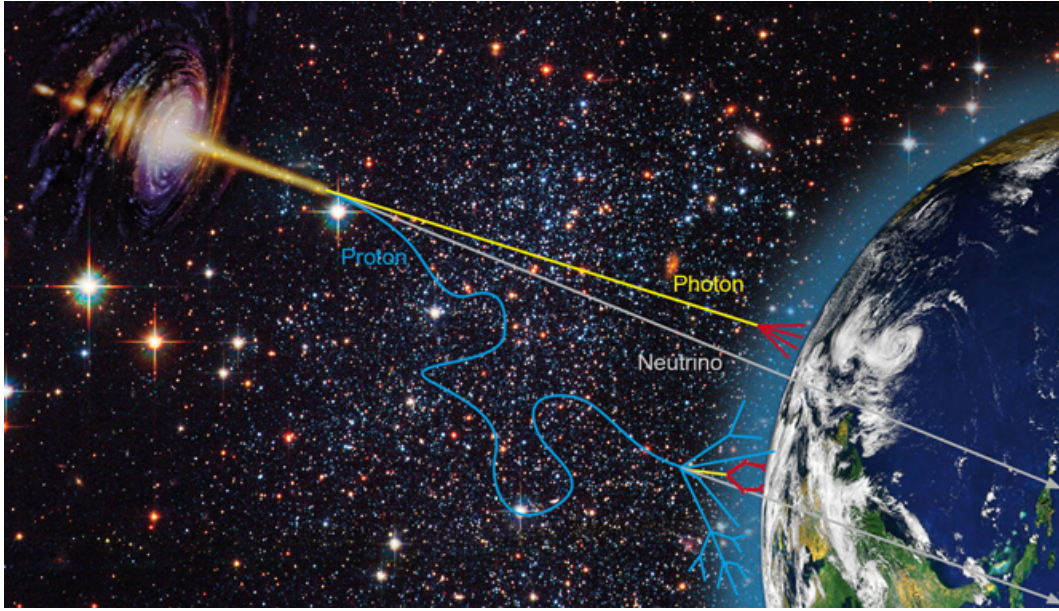


Figure 15: Illustration of the penetrating power of various radiation through stellar matter. Unlike photons, the neutrino can easily pass through solid matter. [23]

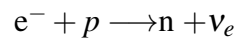
neutrino detectors without interacting. It is therefore only possible to collect useful data when the neutrino flux is high enough that some neutrinos interact with the detecting volume. It follows then that a neutrino detector must be large and run for a long time in order to gather useful data. The astronomical phenomena that can be observed by neutrino detection are fusion reactions in our nearby sun as well as a supernova located within the Milky Way galaxy. Presently, only a few supernova neutrinos have been detected after SN1987A [24].

2.4.1 Supernova Neutrino Production

The final stage of a sufficiently massive star's life is silicon burning, raining iron ash into the core of the star until it inevitably reaches a critical point at which the supernova begins. This critical point is called the Chandrasekhar limit ($\sim 1.4M_{\odot}$) and in this context refers specifically to the mass at which the unburning core of an aged star will either collapse into

a white dwarf or explode as a supernova.

Since relatively little energy is released by the fusion into iron, the core of the star continues to be compressed and heated until electrons and protons are forced to combine, rapidly forming neutrons and neutrinos.



This neutronization is endothermic and therefore accelerates the rate of gravitational compression of the core. The star continues to rapidly contract increasing the core in mass density to a similar scale of the atomic nucleus. The neutronized core is now collapsing with such energy that the particles come within range of the strong force. As a result, a core “bounce” occurs ejecting the core constituents toward the inward falling outer layers of the star.

This initial shockwave slows as the inward falling matter from the outer layers of the star collide with the outward moving matter from the core. The continual release of neutrinos from the core however is energetic enough to heat this mixture until the explosion is resumed. This continues the release of energy of the star in the form of neutrinos as well as produces nuclei heavier than ^{56}Fe .

The neutrinos released during the early supernova process can escape the early supernova shockwave and as such can reach the earth minutes to hours before the early light from the supernova. A Model of the neutrino luminosity spectrum as a function of time can be seen in Figure 16.

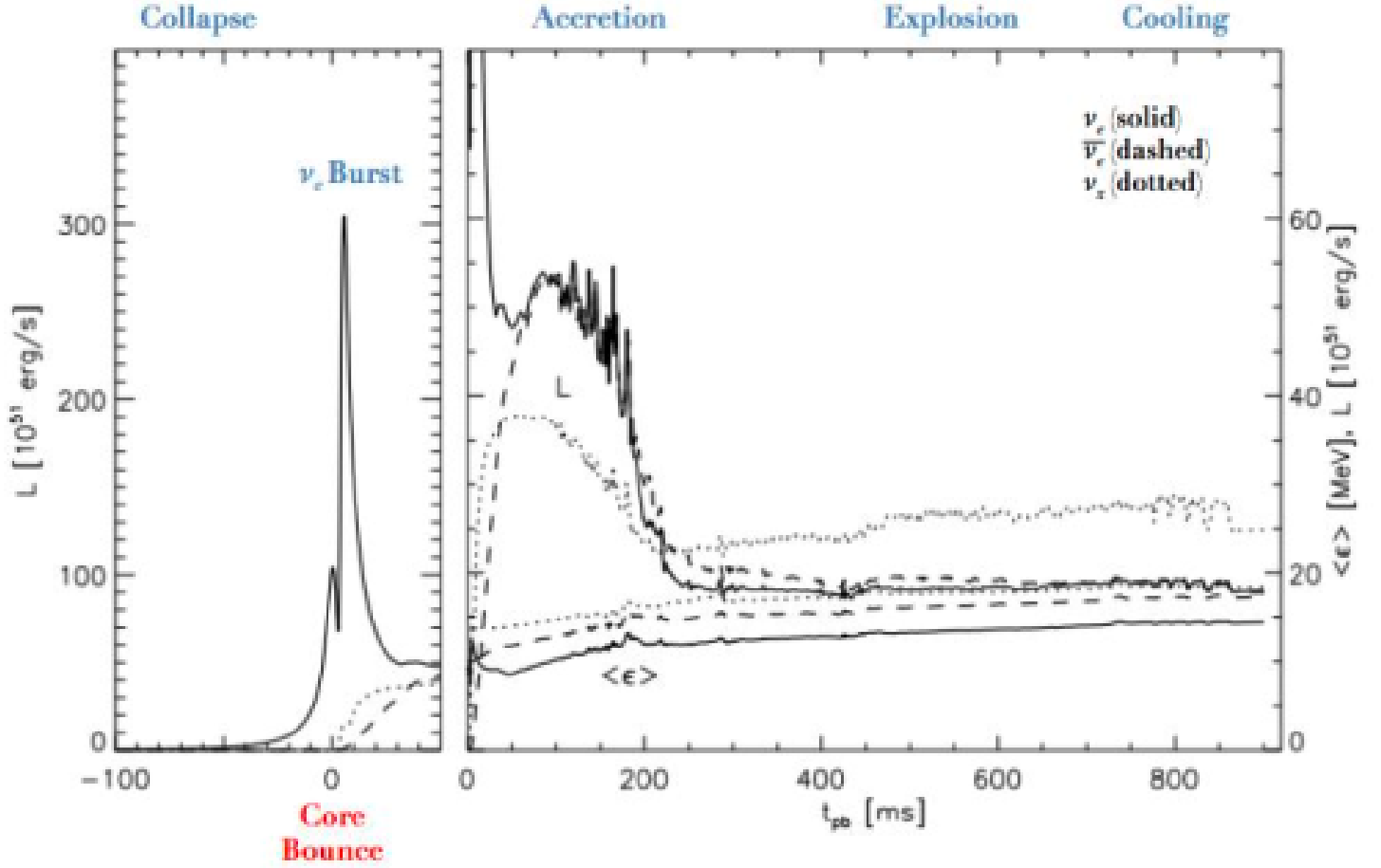


Figure 16: Buras model of supernova core collapse (time-dependent luminosity). The left portion of the graph shows the neutrino luminosity during the core collapse and bounce portion of the supernova on a much larger scale than the right portion since the neutrino luminosity is much greater. During the core bounce phase the electron neutrino is predominantly released. After the bounce and through the course of the supernova explosion electron neutrinos, electron anti-neutrinos as well as tau and mu neutrinos are released in a similar proportion to one another.[\[22\]](#)

2.4.2 Supernova Neutrino Detection

Observation of neutrinos from SN1987A established that neutrinos are heavily involved in the supernova process. It is still useful to strive for further measurement of supernova neutrinos to confirm existing models and learn more about the explosion process itself. [25]

Many neutrino detectors have been built for the study of several processes other than detection of supernova neutrinos. It also happens that many of these same detectors would also be sensitive to a neutrino signal from a supernova. It is therefore not unusual for the detection of supernova neutrinos to be a secondary goal of many neutrino experiments around the world. The variety of neutrino detection mechanisms allows for a wide coverage of different neutrino properties.

The method for detecting neutrinos makes use of the interaction of incoming neutrinos with a target. Typically the result this interaction can be measured, in turn revealing information about the neutrino.

Since neutrinos from supernovae are emitted isotropically, the intensity decreases with the square of the distance between the earth and the neutrino source. It is true then that not only the magnitude of energy released, but also the distance to the supernova is related to the rate of neutrino data production. For this reason, the study of supernova neutrinos is usually limited to galactic supernova.

A large collaboration of many neutrino experiments which are capable of detecting supernova neutrinos, called SNEWS, is continuously monitoring for galactic supernovae to alert astro-physicists of the phenomena as soon as is possible. As the early supernova neutrinos

can arrive at earth minutes to hours before the light signal, the early warning may also precede the arrival of early supernova light by a similar factor.

3 Introduction to SNO+

SNO+ is the successor to the original SNO experiment; a kilo-tonne scale multipurpose liquid scintillator detector designed to observe neutrinoless double beta decay in ^{130}Te , geo and reactor neutrinos, low energy solar neutrinos and other exotic decays in water. Being a large scale neutrino detector, it is sensitive to supernova neutrinos originating within our galaxy and can participate in the SNEWS (SuperNova Early Warning System) multi-detector collaboration.

3.1 Apparatus

SNO+ re-purposes the SNO apparatus with a few modifications for the new experimental goals. The detector is located beneath 2 km of norite rock in the Creighton Mine near Sudbury Ontario in Canada. The rock provides approximately 5890 ± 94 [14] meters water equivalent of shielding reducing the cosmic muon flux from $1.67 \times 10^{-2} \frac{\mu}{\text{cm}^2 \cdot \text{s}}$ at the surface to $3.1 \times 10^{-10} \frac{\mu}{\text{cm}^2 \cdot \text{s}}$ [26] translating to ~ 63 muons per day [14] through the detector volume.

The hosting facility, SNOLAB, is maintained as a class-2000 clean room establishing a baseline for cleanliness which the hosted experiments exploit. This cleanliness level serves to significantly reduce background radiation from uranium and thorium daughters present in the mine dust and allows experiments at the facility to easily reduce these backgrounds even further.

The SNO+ Acrylic Vessel (AV) and photomultiplier tube support structure (PSUP) are suspended from a steel deck structure at the top of a large cavern which acts as the contain-

ment vessel for 7000 tonnes of ultra-pure water shielding (Figure 17). The ultra-pure water is continuously monitored for basic purity with conductivity measurements and periodically monitored for radiopurity with radioactive assay techniques.

The PSUP is a stainless steel geodesic sphere surrounding the AV used as a frame to support approximately 9500 photomultiplier tubes (PMTs) and their individually associated reflective hex cell concentrators resulting in an effective photocathode coverage of 54% [7]. The photomultiplier tubes themselves are Hamamatsu R1408 PMTs which were reused from the original SNO experiment. Each of these is submerged in the ultra-pure water volume and connected individually via an impedance matched cable to the data acquisition system on the deck above. The majority ($\sim 99\%$) of these

PMTs face inward toward the detector volume while a select few face outward for background rejection purposes.

Between the PMT array and the inner detector volume is a 3 m thick region which is also filled with ultra pure water. During normal operation, this region is the primary injection point of the freshly purified ultra-pure water in the outer detector region. Since the flow of

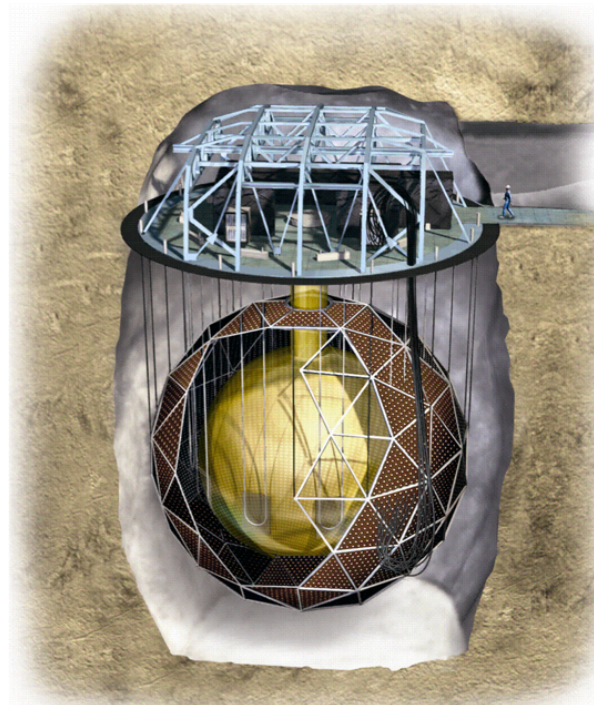


Figure 17: 3D Rendering of the SNO+ detector. Not pictured are the hold down rope net and anchors which are a major modification to accomodate the buoyancy of liquid scintillator in water. [27]

water from inside the PSUP to the outer regions of the cavity is significantly restricted, the region between the PSUP and the AV is considered to be fairly independent and is maintained at a higher level of purity than regions outside of the PSUP.

The target material within the instrumented volume is contained within the 12 m diameter AV. This is the volume which originally contained the heavy water target that was used during the SNO experiment.

Since the original SNO experiment used heavy water as the target contained by the acrylic vessel, there was no need for the AV to be held down as the density of the D_2O would more than counteract the buoyancy of the AV in light water. The lower density of the liquid scintillator target ($\sim 0.84 \frac{kg}{l}$) however makes this configuration problematic for SNO+. This necessitated a major modification to the original design where an additional rope net, shown in Figure 18, had to be installed to hold down the AV.

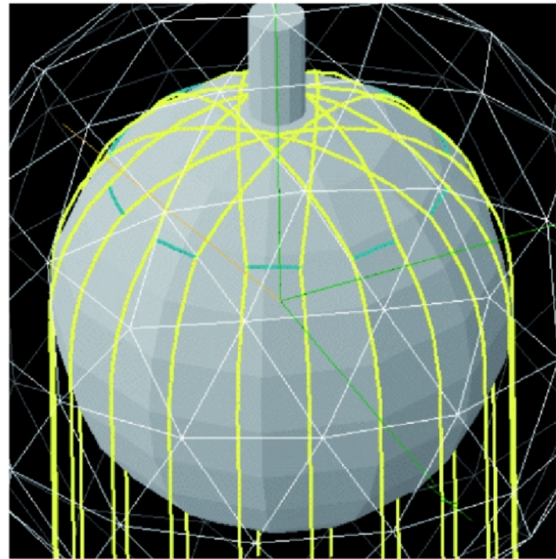


Figure 18: 3D simulated rendering of the orientation of the SNO+ Acrylic Vessel hold down rope net. [14]

The hold down ropes are secured to the floor of the rock cavern with heavy duty anchors secured with rock bolts drilled into the rock. Each anchor was subsequently coated along with the cavity floor with a thick coating to once again act as a containment vessel for the light water shielding. The hold down ropes were stretched in stages before their final connection

to the floor anchor was made with a load-cell.

Further modifications to the detector include the addition of a multi-fiber light injection system and a redesigned detector universal interface. Multiple fiber-optic cables were installed at strategic locations on the photomultiplier support structure to inject light into the detector from a source located on the deck. These light injection points allow for calibration of the detector without the need to insert a physical calibration source into the detector.

The universal interface, a stainless steel enclosed cap used as a source manipulation glove box, was created to facilitate the clean insertion of calibration sources, host a clean inert cover gas from boil off liquid nitrogen and has provisions for interaction with process systems. These modifications help to reduce the contact of any objects with the scintillator target reducing the risk of contamination of the target material.

In summary, these modifications allow the SNO+ detector to host a 12 m diameter buoyant liquid scintillator detecting volume. The detecting volume is surrounded by 9500 PMTs that record the effects of any light producing phenomena within the detector volume.

3.1.1 Liquid Scintillator

The key to the success of SNO+ will be its liquid scintillator target medium. Linear alkylbenzene (LAB) was selected as the scintillating solvent for the SNO+ scintillator cocktail due to its stability over an extended time frame, compatibility with the acrylic that the AV is made from, high purity directly from the manufacturer, high light yield, high optical transparency and linear energy response. The target medium itself is composed mainly of LAB with smaller amounts of a wavelength shifter and ^{130}Te . Once full, the acrylic vessel will

contain 780 tonnes of the LAB based cocktail.

Molecules of LAB have the chemical composition of $C_{18}H_{30}$ and collectively function as both the scintillating reagent and the main solvent in which other entities will be dissolved to create the final scintillator cocktail. Since LAB emits photons in low ultraviolet wavelengths and the Hamamatsu R1408 PMTs are most sensitive to light in the visible spectrum it is necessary to add a wavelength shifter to this mixture. The chosen wavelength shifter is the fluor 2,5-diphenyloxazole or more commonly referred to as PPO. PPO will be dissolved into the cocktail at a concentration of 2 g/l and will use the energy of incident ultraviolet photons from the LAB to re-emit light of a longer wavelength suitable to the operation of the PMTs. [14]

Though the LAB will arrive from the manufacturer in Becancour Quebec with exceptional purity, a liquid scintillator purification plant (shown in Figure 19) has been erected in the underground facility. During the initial construction of the process plant, a campaign of helium leak checking was conducted on all vessels and piping ensuring that any radon ingress during processing would be minimized. Next the entirety of all process systems were thoroughly passivated with hot citric acid before an extensive flushing with ultra-pure water to remove any loose particles or surface contaminants that may have been created during the construction process. The cleanliness of the scintillator plant as well as the short commute of LAB to the site minimize any contamination of the LAB.

The scintillator process plant processes LAB through a primary distillation phase, then water extraction and steam stripping later to further purify the liquid scintillator. The expected purity after these processes would leave concentrations of uranium and thorium at 10^{-17} g/g

and 10^{-18} g/g respectively. [28]

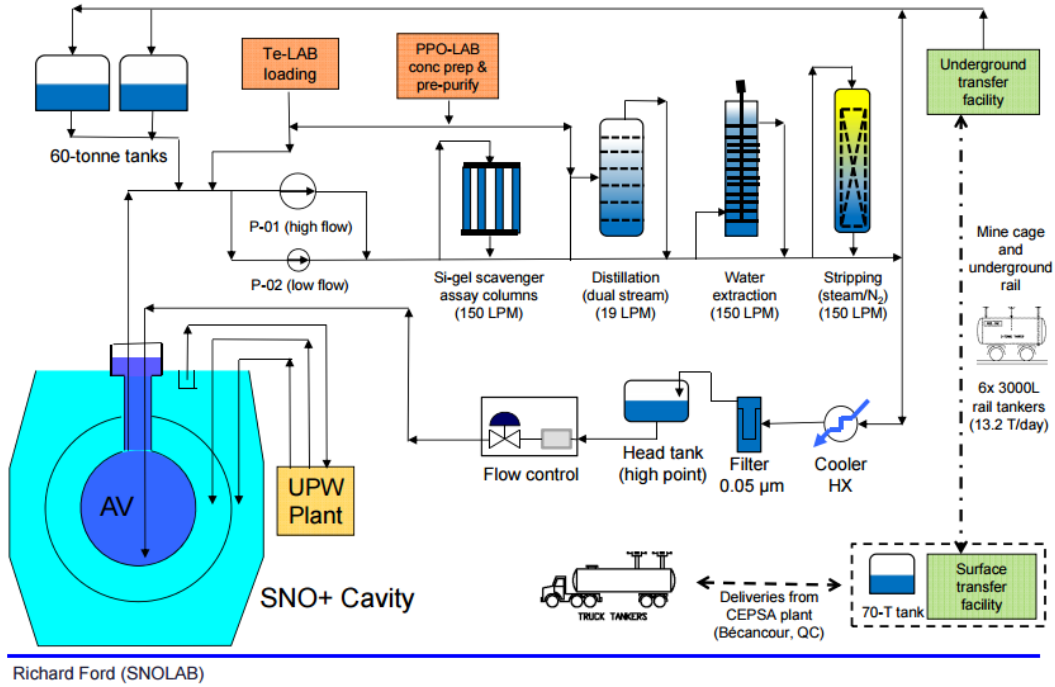


Figure 19: Overview of the scintillator purification process [28]

3.1.2 Detector Electronics

Particles interacting in the detector will produce photons which can in turn be measured by the PMTs. In order to determine the cause and nature the PMT signal, it is important that information with respect to timing and charge be preserved. For SNO, 9500 inward facing Hamamatsu R1408 photomultiplier tubes were specially made to contain exceedingly low amounts of potassium, thorium and uranium. Their pulse rise and fall time is less than 1.7 ns [7]. These same PMTs are now being used to take data for SNO+.

Each photomultiplier tube has an 8 inch diameter photocathode leaving the assembled detector with an overall photocathode coverage of $\sim 30\%$. This was improved upon sig-

nificantly with the addition of a reflective concentrator housing (Figure 20) for each PMT increasing the effective photocathode coverage to 54%. [30]



Figure 20: SNO+ R1408 PMT mounted in a reflective "hex cell" housing housing.[29]

When an inbound photon strikes the photocathode, a photo-electron is released with 25% quantum efficiency. The electron is then accelerated and guided by an electric field towards the first dynode stage. The now energetic electron strikes the dynode releasing a few more electrons through secondary emission. These electrons are then continually focused and accelerated through several

gain stages whose total resultant gain is $\sim 10^7$. The generated negative current pulse can be measured by the detector electronics.

Each cable connecting the PMT to the detector electronics powers the PMT and propagates its signal to the data acquisition system. In order to preserve signal timing, single RG59/U cables of identical lengths are connected to each PMT. The PMTs are then coupled capacitively to the detector data acquisition system. A schematic overview of the analog trigger system can be seen in Figure 21.

The RG59/U cable terminates at the crate end in one of 4 paddle cards mounted to a PMT Interface Card (PMTIC). Each PMTIC provides high voltage to its 4 paddle cards through 4 user operable relays. Since each paddle card is responsible for controlling the voltage supplied to 8 PMTs each of the 8 PMT channels on a paddle card have individually biased

output voltages through the use of high accuracy resistors. This means that an operator has the ability to enable or disable sets of 8 PMTs at a time using a single board mounted relay on the PMTIC controlled by the detector control software.

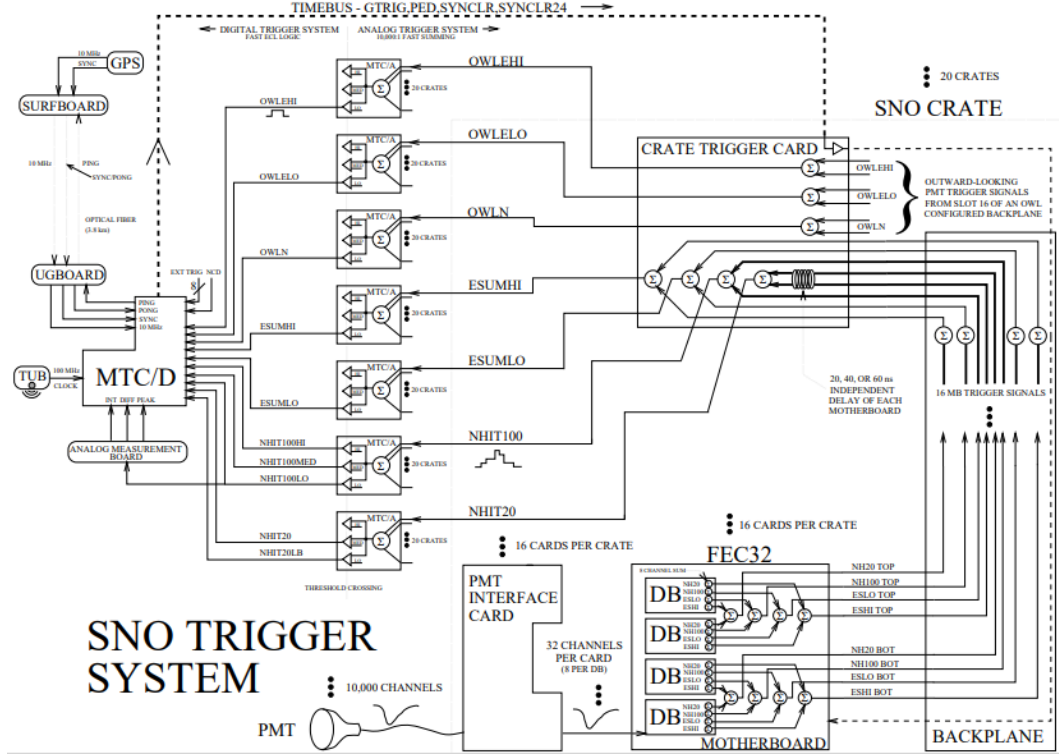


Figure 21: Schematic diagram of the SNO+ detector electronics and trigger system. These are the electronics responsible for monitoring all of the PMTs and triggering the detector based on user specified inputs. [31]

Capacitive coupling of each channel on the PMTIC happens immediately before the signal is passed to one of 4 daughter boards on a Front End Card (FEC). The daughter board produces two logic pulses upon receiving a PMT pulse: one 20 ns long (NHIT20) and one 100 ns long (NHIT100). The daughter board also passes the capacitively coupled raw signal which propagates through an interconnections bus called the back plane along with the NHIT20 and NHIT100 pulses to analog summing logic on the Crate Trigger Card (CTC).

These analog sums are then passed to the Master Trigger Card-Analog (MTCA) which once again creates an analog sum of all of the crate sums. If the sum of the logic pulses at the MTCA is above a user specified threshold for NHIT100 or NHIT20, the MTCA will trigger and pass the signal pulses to the Master Trigger Card Digital (MTCD) which in turn will trigger if the trigger parameters specified by the detector operator are masked in. For this trigger then a Global Trigger IDentification (GTID) number will be issued to all crates. The electronics and data acquisition system then capture all of the data which are then digitized for further processing. This constitutes an "event" which all of the information about the event in the form of channel data and detector state is written to the data file [32].

3.2 Experimental Goals

SNO+ is a multi-purpose neutrino detector with many physics goals to contribute to: neutrinoless double beta decay, low energy solar neutrinos, reactor neutrinos, geo neutrinos, supernova neutrinos and nucleon decay. While SNO+ is a multipurpose detector, that does not mean that all of its experimental interests can be explored simultaneously. With that in mind the detector will be operated in separate phases which progress with the construction and commissioning of the detector.

For the first phase, the detector will be operated as a water Cherenkov neutrino detector sensitive to some solar neutrinos and muons as well as exotic nucleon decays. During this phase, the limit on partial lifetime decay of protons and neutrons in the oxygen nucleus were found to be 3.6×10^{29} y and 2.5×10^{29} y respectively [33]. Limits on dinucleon decay modes were found to be 2.6×10^{28} y and 4.7×10^{28} y for pn and pp modes respec-

tively [33]. A measurement of the ^8B solar neutrino flux was also made and found to be $2.53^{+0.31}_{-0.28}(\text{stat.})^{+0.13}_{-0.10}(\text{syst.}) \times 10^6 \text{ cm}^{-2}\text{s}^{-1}$ [34].

The second phase of the experiment will use a scintillator cocktail made only with PPO and LAB. The transition to this phase will occur concurrently with scintillator filling operations and the purification of scintillator as it arrives from the manufacturer. The liquid scintillator will be added to the top of the acrylic vessel while the ultra-pure water that is inside will slowly be drawn out from the bottom. This transition is currently in progress with LAB filling the neck section of the AV. Once full the second phase of the experiment will begin with an extremely pure detecting medium making it ideal to explore low energy solar neutrino properties. The target internal and external background levels for this phase are 1.6×10^{-17} g/g uranium and 3.5×10^{-13} g/g uranium respectively. The third phase of the experiment will load the scintillator with the isotope ^{130}Te beginning perhaps the most interesting phase of the experiment probing the theorized Majorana property of neutrinos. For this phase the target internal background level required is not as low as the second phase at 3.5×10^{-13} g/g uranium.

3.2.1 Neutrinoless Double Beta Decay

The primary goal of SNO+ is to explore the Dirac or Majorana nature of neutrinos by searching for neutrinoless double β -decay ($0\nu\beta\beta$) in the isotope ^{130}Te . A double β -decay can be thought of as two single β -decays occurring simultaneously for the same parent nucleus. A double β -decay candidate isotope in this context is one where the single β -decay is disallowed energetically over the double β -decay.

The total number of candidate isotopes for this process is clearly limited then and isotope selection must be considered carefully for any $0\nu\beta\beta$ experiment. Some of the double beta decay candidate isotopes are presented in Table 3.

Isotope	Q-Value	$2\nu\beta\beta$ Half Life	Natural Abundance
^{48}Ca	4270 keV	5.8×10^{22} y	0.19 %
^{76}Ge	2040 keV	1.8×10^{21} y	7.8 %
^{82}Se	3000 keV	9.6×10^{19} y	9.2 %
^{100}Mo	3930 keV	7.1×10^{18} y	9.6 %
^{130}Te	2530 keV	7.0×10^{20} y	34.5 %
^{136}Xe	2480 keV	2.2×10^{21} y	8.9 %
^{150}Nd	3370 keV	9.1×10^{18} y	5.6 %

Table 3: Table of double β -decay candidate isotopes. From this list ^{130}Te was selected due to its high natural abundance and would more practically suit the scintillator cocktail due to its physical properties.

^{130}Te was selected as the double β -decay candidate isotope because it has a high natural abundance, it allows for a produced scintillator cocktail to be both removed and re-purified if necessary, and the loading can be easily scaled up to fill the detector. The double β -decay of ^{130}Te to ^{130}Xe would release 2.528 ± 0.013 MeV [35] divided amongst the released electrons and neutrinos. A concentration of approximately 0.5% by weight or 2.3 T of tellurium will be added to the scintillator volume to form the target medium [14]. If neutrinoless double beta decay were observed, the neutrino would have to be capable of being its own anti-particle. This would rule out the Dirac nature of distinct boundaries between anti-particles and particles. This would serve as the first example of the Majorana mass theory which predicts that uncharged fermions could exhibit relativistic wave equations equalling that of the same fermion's anti-particle. In order to explore this process a certain decay isotope must be chosen which can only decay by double beta decay, since neutrinoless double beta decay

would indicate that neutrinos behave according to the Majorana theory rather than Dirac theory.

Neutrino oscillation experiments measuring survival rates of the different flavours of neutrinos provide information about mixing angles and indicated that the flavours had small mass differences between them. The neutrino mass hierarchy can be predicted based on their results. There are two possible neutrino mass hierarchies which have been predicted based on the square mass differences that have been measured during various neutrino oscillation experiments. This required both the results from solar neutrino experiments and atmospheric neutrino experiments since the flavour dynamics of both systems would differ greatly due to the matter enhanced MSW effect present for solar neutrinos. Since the neutrino mass cannot yet

be determined and only the mass difference was measured, there are two proposed mass hierarchies, normal and inverted. The mass splitting values $\Delta m_{sol}^2 = \Delta m_{21}^2$ and $\Delta m_{atm}^2 = \Delta m_{32}^2$ are shown in Figure 23.

Both hierarchies suggest a set of assumptions that can be used to calculate the Majorana mass, $m_{\beta\beta}$, a vital value relating to Majorana neutrino flavour mixing phases. For normal

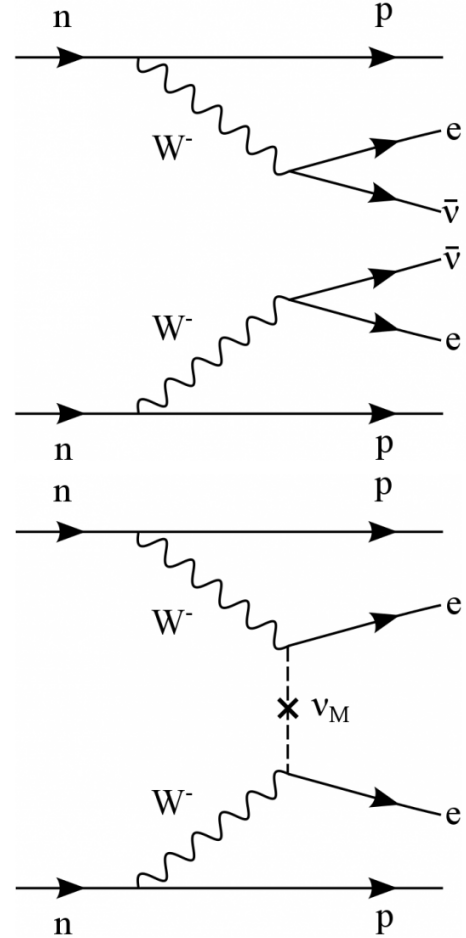


Figure 22: Feynman diagrams showing two neutrino and neutrinoless double beta decay. Top: $2\nu\beta\beta$ Bottom: $0\nu\beta\beta$ [36]

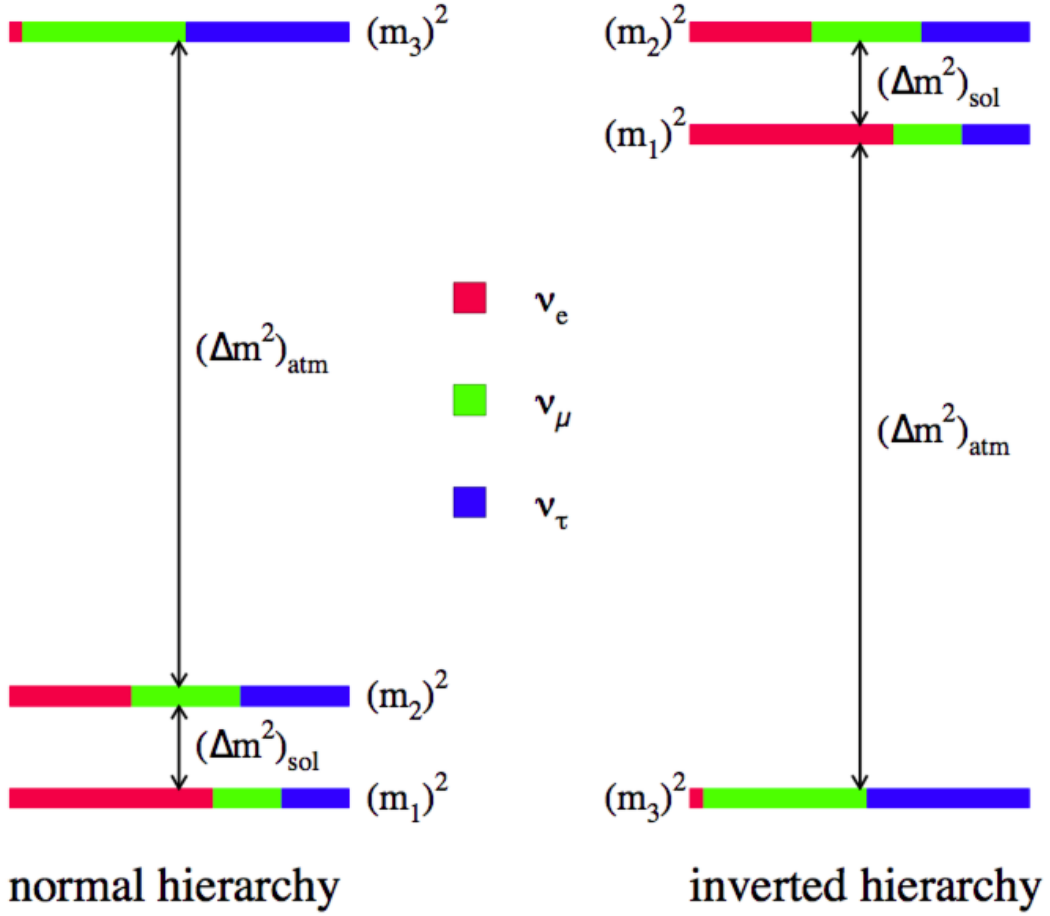


Figure 23: Neutrino mass hierarchies with colours representing the relative sizes of associated mixing matrix elements. Normal hierarchy on the left and inverted on the right. $\Delta m_{\text{sol}}^2 = \Delta m_{21}^2$ and $\Delta m_{\text{atm}}^2 = \Delta m_{32}^2$ are as measured by solar and atmospheric neutrino experiments with the relative proportions of neutrino flavours indicating and overall oscillation phase.[37]

hierarchy the assumption is that the heavy neutrino will have mass approximately equal to the mass difference between it and the light neutrino masses. For an inverted hierarchy the assumption is that the heaviest neutrinos have the same mass and the mass of the lightest neutrino is very small by comparison. These assumptions lead to two different expressions for the effective neutrino mass $\langle m_{\beta\beta} \rangle$ which are for normal and inverted hierarchies respectively:

$$< m\beta\beta > \approx \left| \sin^2 \Theta_{12} \sqrt{\Delta m_{12}^2} + e^{2i\alpha} \sin^2 \Theta_{13} \sqrt{\Delta m_{23}^2} \right| \quad (3.1)$$

$$< m\beta\beta > \approx \sqrt{|\Delta m_{13}^2| (1 - \sin^2 2\Theta_{12} \sin^2 \alpha)} \quad (3.2)$$

[38]

It is expected that the rate of neutrinoless double β -decay would be some small fraction of the rate of all double β -decays for the double β -decay candidate isotope in question. As such, neutrinoless double β -decay experiments seek to observe or to set lower limits on the half-life of neutrinoless double β -decay. The neutrinoless double β -decay rate is given by Equation 3.3.

$$^{0\nu}T_{1/2}^{-1} = < m\beta\beta >^2 |M^{0\nu}| G^{0\nu}(Q, Z) \quad [38] \quad (3.3)$$

where $G^{0\nu}$ is a well known constant, $M^{0\nu}$ can be calculated to some extent but depends highly on assumptions and $< m\beta\beta >^2$ is the effective majorana mass given by:

$$< m\beta\beta >^2 = \left| \sum_i U_{ei}^2 m_i \right|^2 \quad (3.4)$$

where U_{ei}^2 are neutrino mixing matrix elements involving both known mixing angles and unknown Majorana mixing angles.

The lack of measurement of the Majorana phases prompts the calculation of the range of possible Majorana masses as a function of the lightest neutrino mass which is of course dependent on the neutrino mass hierarchy. [38]

Since the Majorana phase (α in Equation 3.2), is currently unknown, the inverted hierar-

chy equation could have a cancellation between the two terms leaving an effective neutrino mass of zero. On the other hand, there can be no canceling in the inverted hierarchy effective mass equation resulting in a minimum or maximum value for the effective neutrino mass depending on α . The possible effective neutrino masses can be plotted to show the ranges which might be expected depending on the Majorana phase. Currently, $0\nu\beta\beta$ experiments can probe effective neutrino masses for a part of the inverted mass hierarchy region (Figure 24) with the intent to push deeper and deeper toward the normal hierarchy region.

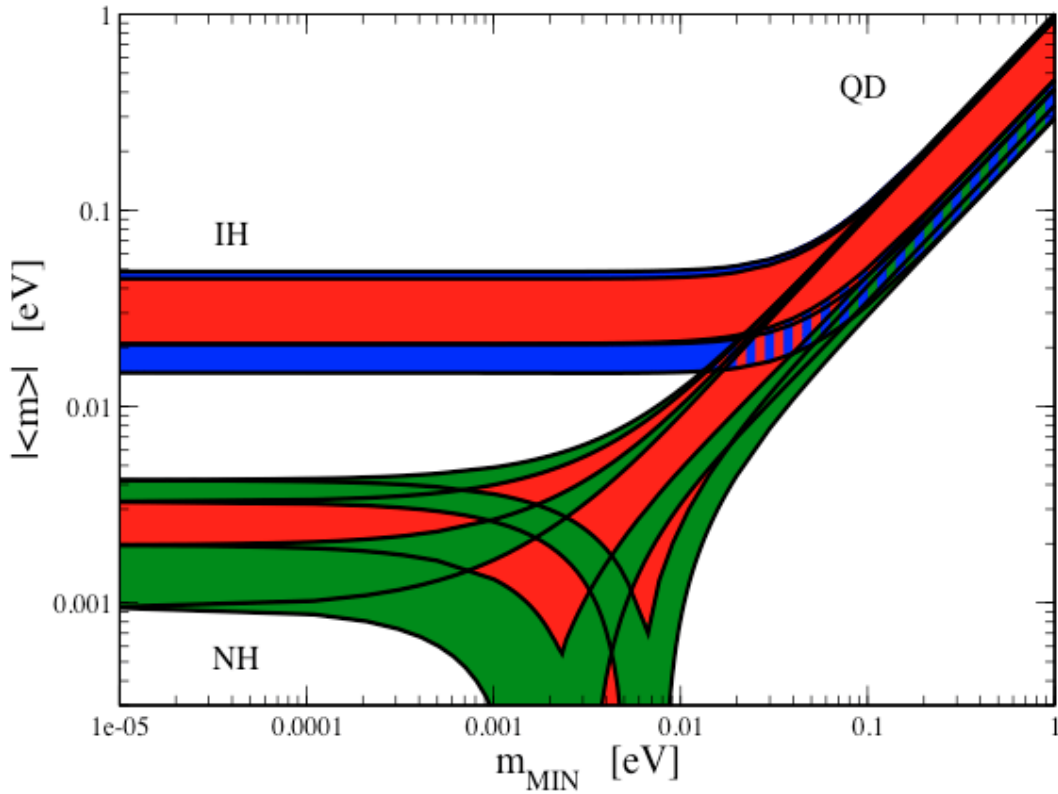


Figure 24: Effective neutrino mass as a function of the lightest neutrino mass. Width of the bands are determined by the Majorana phase factor. NH is normal hierarchy and IH is inverted hierarchy. Overlapping regions are degenerate. Red blue and green bands correspond to values obtained allowing charge parity violation and for 1σ variation in mixing parameters. [39]

With ^{130}Te as the double beta decay candidate isotope, SNO+ is poised to set world lead-

ing limits on the measurement of the neutrinoless double beta decay half-life. The effective mass limit that SNO+ will be able to set is projected to be between 38-95 meV [14].

The sensitivity of SNO+ to the $0\nu\beta\beta$ signal is constrained by the radioactive backgrounds shown in Figure 25, the most prevalent and irreducible of which is the $2\nu\beta\beta$ background. Reduction of the other backgrounds will serve to increase the sensitivity of the experiment.

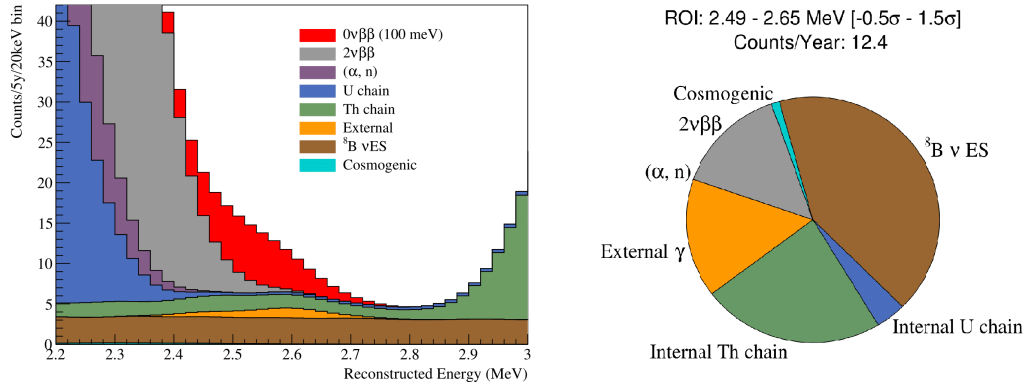


Figure 25: SNO+ Backgrounds and $0\nu\beta\beta$ signal. ^8B solar neutrinos are by far the largest source of background events and are irreducible. Reducing backgrounds from other sources will still increase $0\nu\beta\beta$ sensitivity of SNO+. On the left we see the stacked background count spectrum in the range of (2.2-3.0 MeV) over 5 years. The pie chart on the right shows the annual relative backgrounds for $0\nu\beta\beta$ in the range of 2.49-2.65 MeV. [40] [41]

3.2.2 Solar Neutrinos

SNO+ will be sensitive to neutrinos created during nuclear fusion in the sun. Of particular interest are neutrinos produced from the pep reaction ($p + e^- + p \rightarrow {}^2\text{H} + \nu_e$) and neutrinos produced by the CNO cycle. The measurement of pep neutrinos will reveal a greater understanding of neutrino interactions within dense matter. Measurement of CNO neutrinos, which make up less than 2% of the total emitted solar neutrino flux, require that backgrounds

be very low so that they can be distinguished from the pp solar neutrino flux. This measurement however will provide information about the metallicity of our sun. SNO+ will be capable of making both of these measurements at world leading levels due to the extremely low cosmogenic and radioactive backgrounds of the apparatus. These measurements would lead to further development of the solar model which was explained in Chapter 2.

3.2.3 Reactor Neutrinos

The process of nuclear fission produces unstable matter which subsequently decays emitting anti-neutrinos in all directions. All nuclear reactors produce a neutrino flux however SNO+ is located within close proximity to three nuclear reactors of particular interest: Bruce nuclear generating station, Pickering nuclear generating station and Darlington nuclear generating station which are expected to be the production sites of 60% of the total reactor neutrino signal at SNO+ per year. It is expected that approximately 90 reactor anti-neutrinos will be detected every year. The relative distances to SNO+ of these three reactors coincides with the matter enhanced MSW neutrino flavour oscillations in matter. This can be explored during all phases of the experiment and will allow for a new measurement to be made for the Δm^2 neutrino oscillation parameter.

3.2.4 Geo Neutrinos

It is thought that similar to the core of the sun, much of the dynamics of the inner layers of the earth is driven by nuclear processes. SNO+ will be able to make a direct measurement of neutrinos released during any of these processes due to the great depth and uniformity of the

surrounding rock. SNO+ will be able to set limits on the radioactivity and concentration of uranium and thorium in the surrounding crust since these primarily and regularly β^- -decay releasing electron type antineutrinos. This will aid in creating a map of the distribution of radioactive elements in the earth's crust and is the first time this measurement will be made in the Canadian shield. Detection of geo antineutrinos will occur through all phases of SNO+.

3.2.5 Supernova Neutrinos

SNO+ will be sensitive to the neutrino signal from a galactic supernova for the duration of the experiment's life. A galactic core collapse Type II supernova will release the bulk of its gravitational binding energy in the form of neutrinos over the course of a few seconds. This extremely large neutrino signal would undoubtedly create massive amounts of data that may reveal more about the supernova process and the formation of heavy elements in the universe. In preparation for a supernova signal, SNO+ implements a few key mechanisms to manage the bursts of neutrino events associated with a neutrino flux from a galactic supernova. The primary focus of these mechanisms is to allow for retention of high rate data in a burst file and prompt notification of supernova like events. The focus of this thesis is a statistical analysis of burst files which were generated during the commissioning and early running of SNO+ with the goal of aiding in SNO+ preparedness for a supernova neutrino signal.

4 Detector Response

SNO+ is sensitive to minuscule amounts of radiation. This is by design to ensure that low energy neutrino signals can be processed into useful physics data. This however presents the problem that high rate light signals could overwhelm the detector. These signals could be generated from some issue with the apparatus or from a real physics event including those produced by a supernova. This means that the detector must be able to dynamically respond to a multitude of signal types and manage the data appropriately in order for meaningful analysis to be conducted.

Part of the control strategy involves a detector operator whose job it is to monitor and respond to the system. Monitoring and response is achieved through the use of three programs: ORCA (Object-oriented Real-time Control and Acquisition)[40], XsnoED (event display)[40] and the SLOW controls (monitoring of parameters that are not time critical)[40].

The detector is controlled through ORCA. Here detector trigger thresholds are set based on the user specified detector operating mode. Ideally the detector would be running with physics trigger thresholds all the time but since occasionally calibrations are run, work is done near the detector and malfunctions do occur, it is necessary to have other run types.

XsnoED, shown in Figures 26 and 27, serves as the graphical representation of various parameters of the detector. Here the operator will generally have a live representation of events occurring in the detector with traits of a given pulse represented by a coloured scale following the geometry of the detector or of crate space. There are other parameters that can be viewed with XsnoED such as waveforms but these are generally not the detector operator's

primary focus when monitoring the detector.

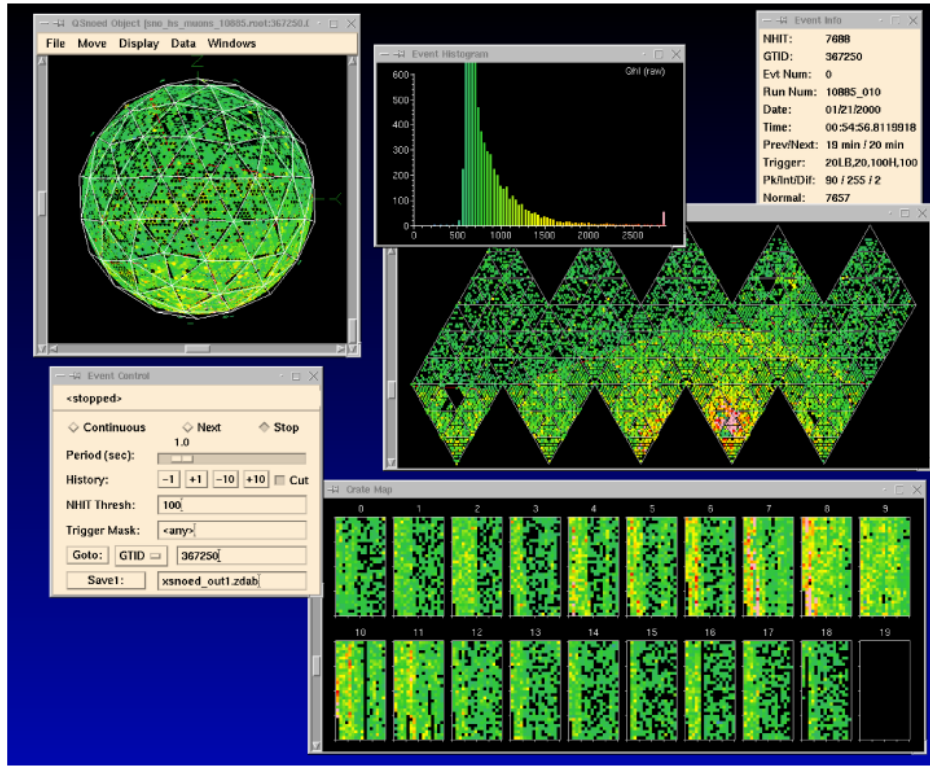


Figure 26: A large event displayed in detector view and in crate space with XsnoED. This particular view represents charge information from the event. Channels higher in charge appear red/pink with low charge channels appearing green. Intermediately charged channels are coloured on the intuitive gradient based on these end points. Each of the 19 crates can be seen in the lower right of the picture. A 3-D view of the PMT array can be seen along with the flattened 2-D view of the PMT array in the center right. A histogram of the same information is also displayed showing the distribution of the number of channels showing a specific charge. [32]

The slow controls (Figure 28) are monitored with access to the on-line database called couchdb. Here the operator can view information about the lab environment such as deck humidity and temperature, detector water levels, detector differential pressures, operating mode of the water systems, temperature of the water surrounding the detector as well as other parameters. These tools allow the detector operator or experts to easily discern any

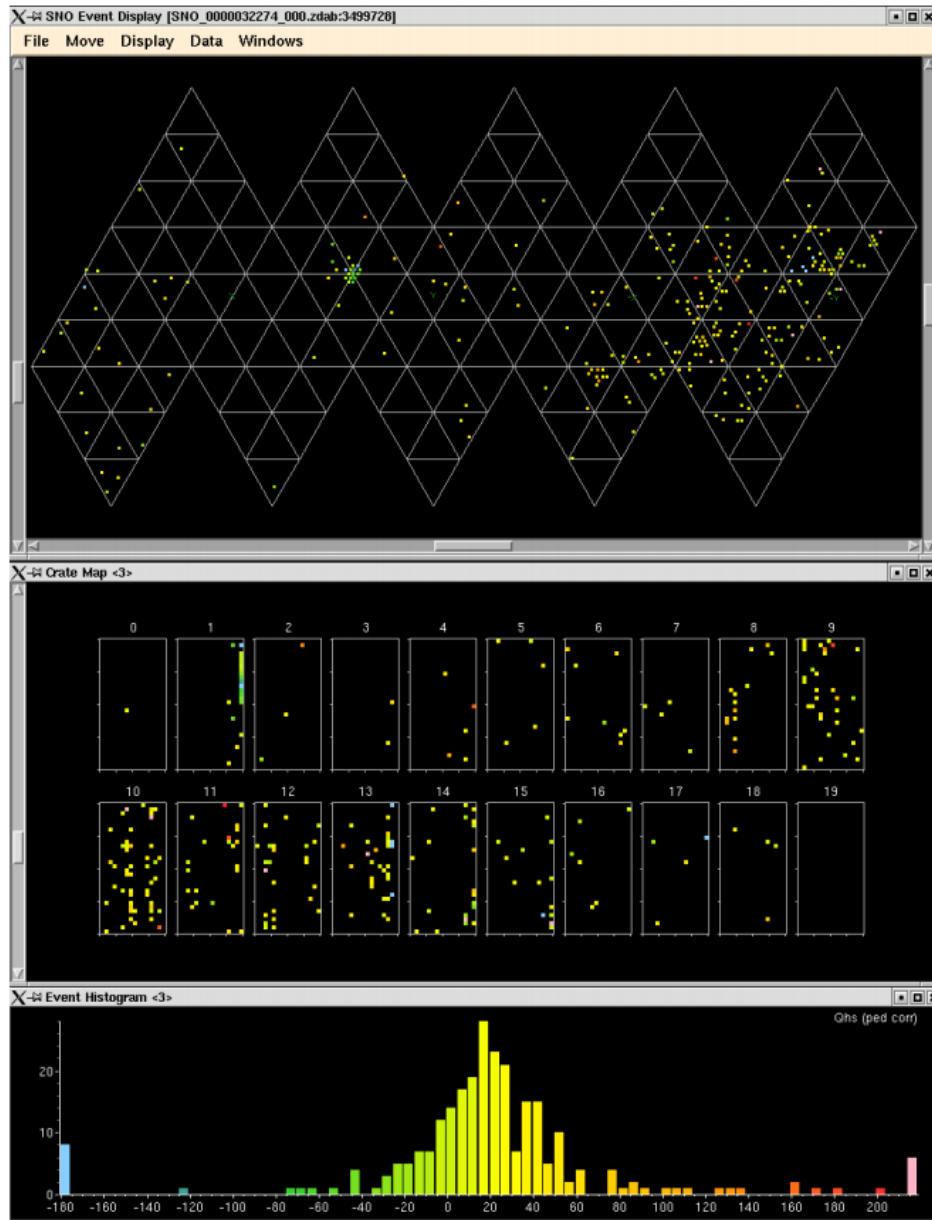


Figure 27: A Flasher event displayed in the XsnoED event display. A flasher occurs when there is a light discharge from one PMT (generally not desired). This light induces pulses in neighboring PMTs before casting light on the opposite side of the detector resulting in a more diffused light pattern. This is evidenced by the tight cluster of hits in a localised area in the detector and on a single card in a crate as well as the diffused light pattern shown by physically distant channels. [32]

issues that may be occurring with the detector environment.



Figure 28: Screen shot of the Slow Controls monitoring webpage

4.1 Analog Signal Chain

The SNO+ detector consists of a large acrylic sphere supported from the top and secured from the bottom with ropes (Figure 29). It is surrounded by a spherical geodesic structure supporting ~ 9500 vacuum photomultiplier tubes which is then completely submerged in water. Each of the triangular portions of the outer steel structure support an array of hex cells - a name given to a PMT and reflector housing assembly (Figure 31). The Hamamatsu R1408 PMTs have 9 dynode stages, a focusing grid and a custom water-proof electronics enclosure allowing them to be fully submerged in water (Figure 30). When an incident photon strikes the photocathode of a single PMT, an exponential shower of electrons ensues from one dynode stage to the next until a detectable negative pulse propagates to the high voltage supply. This pulse is ultimately the signal which is processed into a data file by the detector electronics, builder and data acquisition system (DAQ).

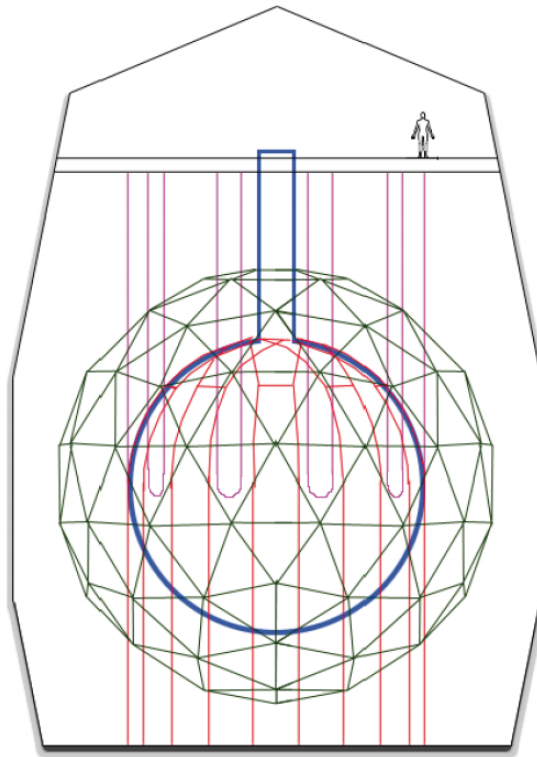


Figure 29: Overall detector construction geometry. Image shows location of AV, AV hold down ropes, AV hold up ropes and PSUP with relation to the deck above. [32]

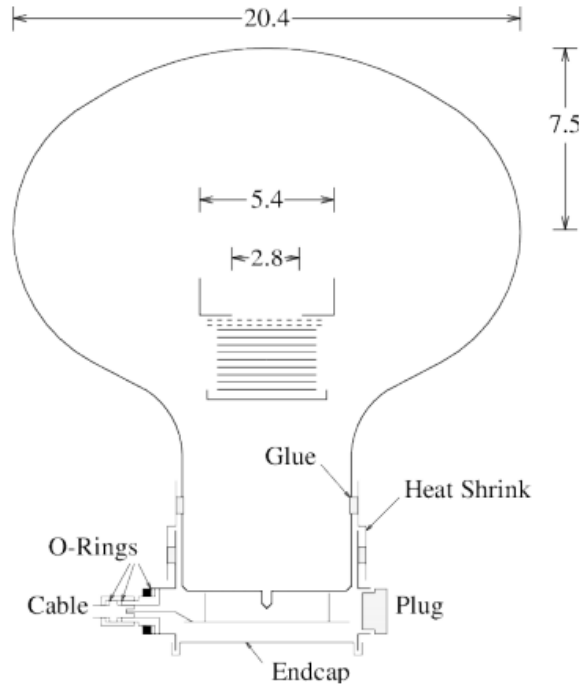


Figure 30: A schematic of the Hamamatsu R1408 PMT and its waterproof end cap. The 9 dynodes are shown as solid horizontal lines. The focusing grid is shown as the dashed lines above the dynodes. The volume inside the end cap is filled with soft silicone gel. The end cap is flexible to allow for thermal expansion. The plug is sealed with heat shrink and thermal adhesive. Dimensions are in cm. [32]

When the pulse from any number of the PMTs reaches the front end electronics, three logic pulses are created within the electronics which then go through an analog summing regime. These three pulses are called the NHIT20, NHIT100 and the ESUM which are a 20 ns wide square pulse, a 100 ns wide square pulse and a summation of the PMT pulses respectively. If the collective peak of one of the pulses is above a specified threshold the detector will trigger, storing all of the detector data as an event. Two other triggers are formed from a smaller set of PMTs which are the OWLE and OWLN triggers and represent the NHIT and ESUM signals from the outward looking PMTs.

Alongside the physics triggers there is functionality to force the detector to trigger with

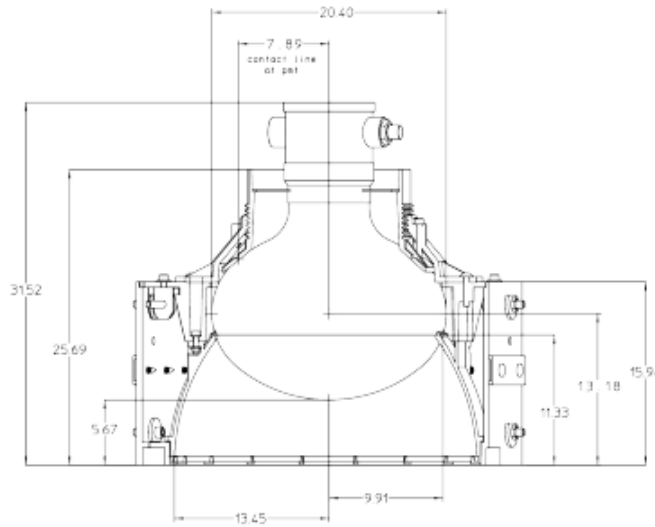


Figure 31: Photomultiplier mounted in a hexagonal ABS cell, which also forms the light concentrator housing. Dimensions are in cm. [32]

external sources. This signal is used to synchronize event timing, identify events and to synchronize or tag calibration events in the detector.

The MTCA (Master Trigger Card Analog) monitors up to three different discriminators constituting different parameters on which detector can trigger (Figure 32). In the event that a certain trigger discriminator is reached by the PMT signal, the signal is passed to the MTCD (Master Trigger Card Digital) where it is compared with a list of specific trigger settings that are specified with the ORCA detector control software.

The specifics of the trigger system and its interaction with the PMTs and DAQ will be discussed in further detail in Chapter 5.

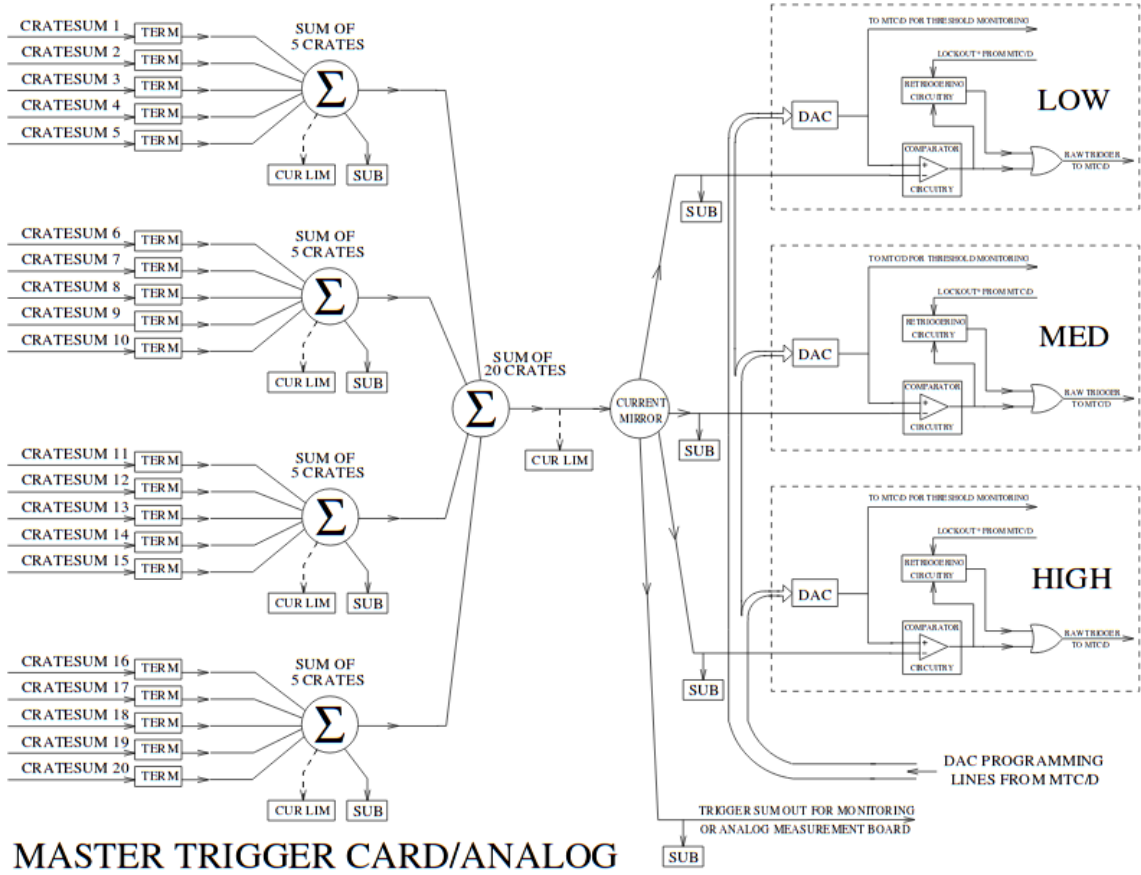


Figure 32: Block Diagram schematic showing the function of the MTCA. Of note are the three stages of discriminators which can be set with the detector control software. [31]

4.2 Digital Signal Chain

Three analog outputs from the MTCA are fed through ECL comparators (emitter-coupled logic comparators are solid state devices which produce an output based on the comparison of two input voltages) with 10V of dynamic range. This dynamic range corresponds to approximately 320 NHIT at 30 mV/NHIT (see Figure 33). Higher thresholds can be set by changing a pull-up resistor resulting in different potential/NHIT ratios feeding the comparators. When the analog sum voltage exceeds a certain threshold the comparator triggers a 20 ns long pulse to the Master Trigger Card Digital (MTCD, Figure 33). This pulse is non-retriggerable and

will not be reset until the signal transitions from high-to-low. The thresholds of the ECL comparators are set with digital to analog converters which are in turn controlled through the detector control software.

The MTCD input receives trigger pulses and controls whether or not the detector is triggered based on those inputs while simultaneously keeping track of the detector timing. The MTCD counts and synchronizes event IDs and stores all relevant event information in memory.

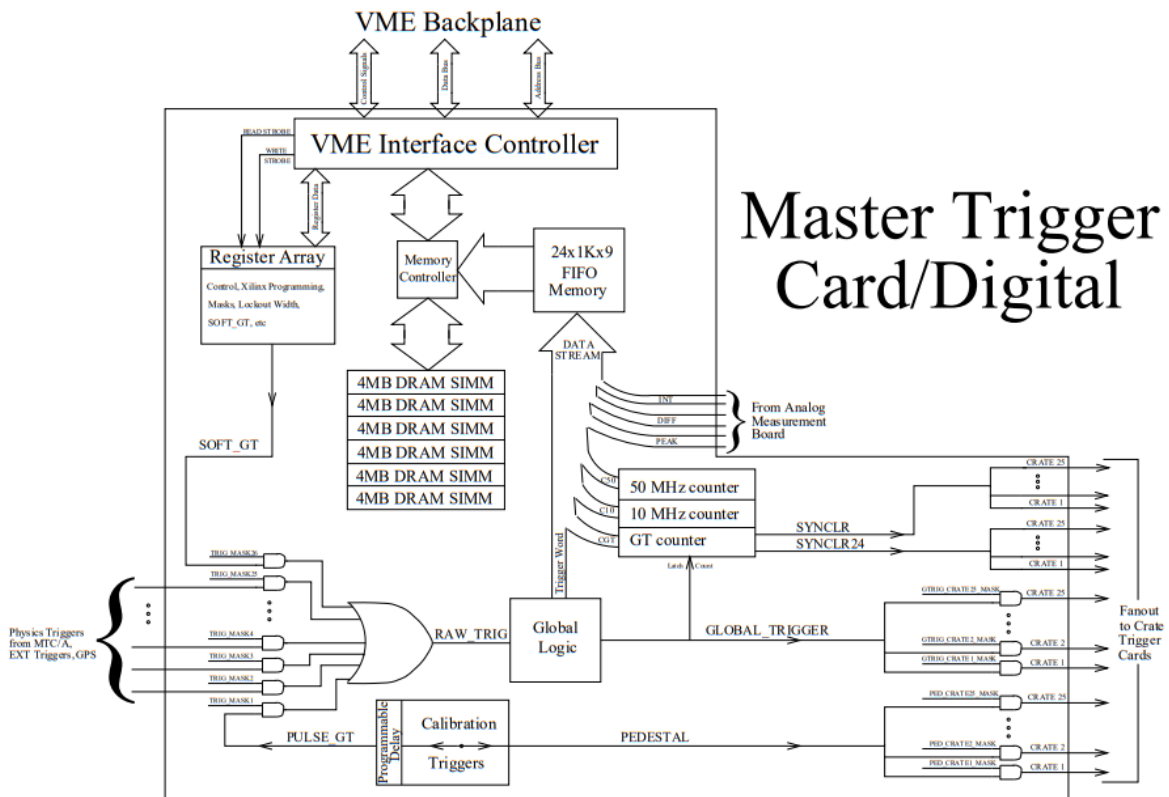


Figure 33: Block Diagram schematic showing the function of the MTCD.[31]

A global trigger is propagated from the MTCD when an input trigger source crosses a user specified threshold (usually is a digital pulse) and that specific condition is "masked in" to the

current run. The trigger candidates can be created both by the MTCA as well as other external trigger sources such as those used for calibration sources. When a rising edge of a masked in input trigger pulse arrives at the MTCD a "raw trigger" pulse is propagated from the MTCD. This raw trigger creates a global trigger and a 20 ns long lockout pulse synchronized to the 50 MHz clock on the MTCD. Since the firing of a global trigger signifies an event and events can occur in any part of the 50 MHz clock cycle, this means that events in quick succession can occur at any time less than 20 ns between events. If conditions to trigger another event were to occur before the 20 ns window of the 50 MHz clock expired, the lockout pulse would prevent a new event from being created. The specific type of detector running will decide exactly which specific trigger or combination of triggers causes the global trigger to fire.

The MTCD clocks are frequently synced with a GPS system that communicates between a surface and underground transceiver. The transceivers frequently communicate with each other in order to remove any latency between the units which would result in inaccuracy of absolute time. For supernova neutrino events it is very important that the real time of the SN events is known so that it can be compared with the results of other neutrino detectors worldwide.

The specifics of the digital trigger system and its interaction with the PMTs and DAQ will also be discussed in further detail in Chapter 5.

4.3 Detector Control and Event Builder

ORCA is the user interface through which the detector hardware manipulation and data acquisition can occur. It stands for Object-oriented Real-time Control and Acquisition. It runs

on any Mac OS-X computer that is part of the SNO+ computing network. A detector operator would be familiar with ORCA as the graphical user interface through which the operator can change the run state and the hardware state when they are on shift. Different levels of permission are granted to different operators depending on their familiarity with the detector and its operation. It is possible that the detector hardware could be damaged and so unrestricted access to ORCA's features is only granted to those who are deemed to be "experts". It is not unusual for an operator to contact an on-call expert to help resolve issues with the detector.

The SNO+ builder is responsible for assembling the digitized data and writing it to file for online monitoring. The data itself is comprised of raw PMT, MTCA/D, CAEN digitizer and ORCA data. With this the builder can assemble events together that will later be concatenated into a single ".zdab" file corresponding to the run. The .zdab file will be named after the run during which it was assembled and will contain all of the events which were built by the builder during said run. The exception to this is a burst file which is assembled separately from run files when certain conditions are met and will be addressed in section 4.4.

The four types of data are read into the builder through independent data threads before being compiled into the event file for the data set. ORCA data is read in via ethernet and is written to a circular buffer for processing by the builder. The MTCA/D and FEC data (PMT bundles) are written to another buffer while headers are written to yet another buffer. MTCA/D data and FEC data are built into events on the basis of their GTID. A list of data stream records is maintained and inserted at the proper position in the data stream while events are written to disk. The metadata associated with events in the data stream are then used to label a set of data with the run header identifying the data set and the specific characteristics of

running when that data was taken. The data is written to disk and can be examined for offline analysis.

It is critical that all of these systems be capable of handling large spikes in the data rate as this is what would be expected in the event of a nearby supernova. As mentioned in chapter 1 section 4, a supernova calibration source was created to stress test these systems and ensure that they could handle such a high event rate.

4.4 Burst File Production

The span of run data files is automated to be 1 hour long and to contain all of the detector data in that hour. After that time a new run is started for the sake of clearing the buffers and parsing out the data. During a run any number of events could occur which would cause the detector to trigger. There is typically a baseline rate at which the detector will trigger creating events that are mostly noise. In the midst of this, physics processes can occur in the detector's target media resulting in real physics data being created and written to file. Runs which are less than 30 minutes long are not analysed.

Sometimes physical processes can cause chain reactions or occur with close enough proximity to each other that the rate at which events are produced will increase sharply for some period of time. If the event rate is sufficiently high for a specified duration, the process is considered to be a "burst" which can contain interesting data including physics data. Currently the burst parameter is defined as 30 events of n_{hit} at least 40 within 10 seconds, ending when rate drops below 10 Hz (averaged over 1 second).

A second smaller data file is created in parallel with the run data that contains only data

from a "burst" in the detector. This file is called a burst file and is produced by a program that is run "nearline" between the data building and the writing out of data files. This is done by a software trigger known as the Level II (L2) trigger whose main intended purpose is to allow the detector to run at a low hardware threshold for supernova detection without compromising low energy data retention permanently. A supernova burst could be identified from other bursts by the timing and charge profiles of events as they occur in the detector with time, and the nature of the hit geometry as we would expect to have a rapid succession of neutrino events which follow the neutrino luminosity profile indicated earlier in Figure 16 (section 2.4.1).

The L2 trigger system is composed of utilities that manipulate the .zdab files in a program known as Stonehenge. Stonehenge is responsible for the basis of the L2 trigger as well as the associated file management that comes with the parallel data processing. A single master script with a user interface controls a few other scripts which do the actual work of the L2 trigger system.

Data from the event builder is read into the L2 trigger system as its sole input. The Stonehenge script then runs with this data from the event builder accomplishing a few different tasks:

- creating the L1 data file, a reduced data file which is monitored nearline for supernova implications
- writing out the "burst" file - a full version of the burst data to be analysed offline
- production of the final L2 file
- data monitoring

As the data comes off of the event builder, Stonehenge monitors for bursts, records event rates and performs simple cuts allowing for bursts to be analysed separately from the main SNO+ Physics data (Figure 34).

A similar supernova trigger system was employed during the SNO experiment but has evolved with the newer iteration. The main direction of changes that were made for SNO+ were in order to ease analysis of the data.

To reduce the chance of unintended biasing of burst data, the data is only cut on the basis of nhit alone. In that regard, there is no filtering of data from other effects such as the intrinsic background produced by the AV.

The configuration of the L2 trigger is dependent on 12 parameters:

- nhithi - The L2 trigger threshold for physics events
- nhitlo - The special lowered threshold for after large events
- lothresh - Threshold at which to lower the threshold
- lowindow - Time for which threshold is lowered
- nhitretrig - Threshold for saving re-trigger events
- retrigwindow - Time window within which an event is defined as a re-trigger
- nhitburst - This is the minimum nhit to be included in a burst file
- burstwindow - Integration window for identifying bursts
- burstsize - Number of events to trigger a burst
- endrate - Rate to declare a burst ended
- bitmask - This determines which events should be saved regardless of nhit

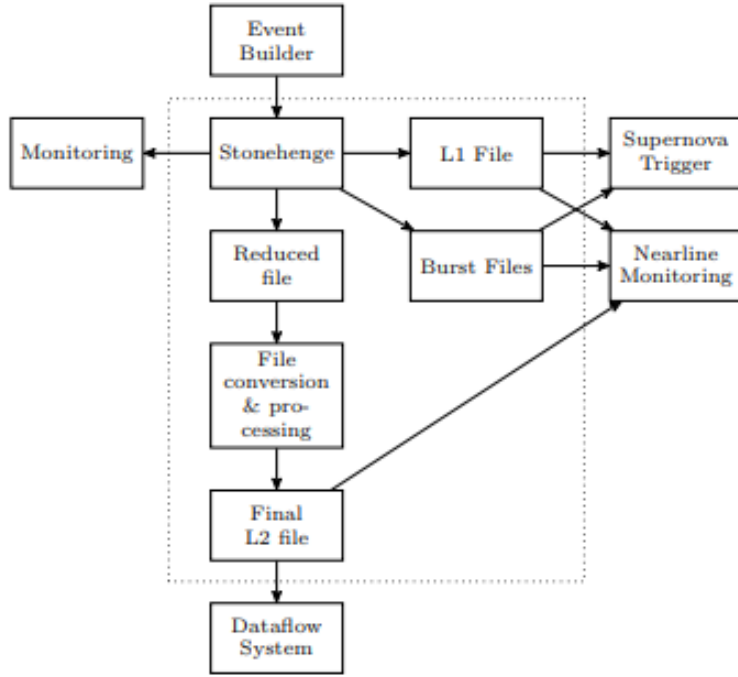


Figure 34: Block diagram showing the data handling of the L2 trigger.[42]

These parameters are read in by the L2 burst trigger and determine the type of data which is contained in the various output branches of the L2 trigger. The exact configuration of this data is selected based on the Run HeaDeR (RHDR) read out from the builder.

The algorithm begins with an event that occurs which is larger than n_{hitcut} . This event is then stored in the burst buffer until a new event is stored in the burst buffer. At this point events older than $burstlength$ are dropped from the buffer and a check of how many events remain in the buffer is performed. If more events are stored in the buffer than specified by $burstsize$, a burstfile is opened. While the burst is ongoing events are removed from the buffer and added to the file if they are older than one endwindow. The number of events in the buffer is once again checked and if it is less than $endrate$ the burst is declared closed and the rest

of the events in the buffer are written out to file. Often times several burst files are created successively as the event rate in the detector fluctuates.

A note on the handling of "orphan" events: Orphans are identified on the basis of their timestamp by the builder. Orphan events have a zero timestamp and usually occur due to timing errors with the detector. They are infrequent but do occur and require that the DAQ software be able to process them somehow. This can complicate burst processing since events will not necessarily arrive at the L2 trigger in order. Orphans are a particular source of these timing issues since they typically occur due to de-synchronization of clocks on the detector hardware. The L2 trigger replaces the zero timestamp of orphans with the timestamp of the event which preceded their entry into the L2 trigger. The same applies for events whose timestamps are less than the previous "good" timestamp received by the L2 trigger. This only applies the timestamps used by the L2 algorithm and not to the timestamp recorded in the event data itself.

5 Burstfile Analysis

Burstfiles are produced by the L2 trigger and then stored on a local server where they can be accessed for offline analysis. They are not currently backed up to Westgrid where they could be processed through grid computing. Eventually this connection should be made so that analysis scripts could be run batchwise to quickly process the data.

Since the burstfiles are direct compilations of events created by the event builder within a burst, they contain all of the same types of data that are stored in the regular data files which are uploaded to Westgrid. It would be possible to narrow down specific events in physics run data that are bursts but this would introduce an unnecessary level of complexity that is taken care of by the L2 trigger. We can take advantage of the data format and employ the same data analysis framework which is used to process the regular SNO+ physics data.

A version of the RAT (Reactor Analysis Tool) [43] framework has been adapted to suit the SNO+ experiment. It has been customized to meet the needs of the collaboration by simulating the materials and geometry of the detector as well as the specific particle/background interactions we would expect to occur in the detector. The SNO+ RAT is the main processor of detector data that reads in detector data files and reconstructs the data for the specific detector state at the time the data was taken before creating an output ROOT file on which data analysis scripts can be run. RAT can also be used without detector data as the SNO+ tailored RAT can also simulate detector data starting from the physics interaction through to the data acquisition. Output data from simulations are once again stored as ROOT files for easy analysis.

The CERN ROOT data analysis framework usually is packaged with the SNO+ RAT. ROOT is typically used to create plots based on data. In the case of SNO+ the RAT output ROOT file contains many branches from which various areas of interest arise. ROOT can also be used as a part of larger scripts and can use discrete or continuous parameters inside the ROOT file in order to determine how the specific data should be treated.

The analysis which is conducted in this thesis first applies the data cleaning algorithms contained in the SNO+ RAT framework to the burst data. Next a vector is created from the data cleaning flags assigned by the data cleaning protocols to each burst file. A statistically driven linear basis transformation is then performed using the vectors as data points. The transformation allows a 2 dimensional representation of the data to be created to facilitate pattern recognition. The specific process implemented to carry out the transformation and represent the data is called principal component analysis.

5.1 Data Cleaning Flags

There are some anomalous signals that can be produced by SNO+ hardware resulting in events being written to data that are not physics events. In fact, like most low background neutrino detectors, most of the stored SNO+ data is not created due to interesting physics processes. As such, these events need to be flagged so that they are not mistaken for background or physics events. Most of these instrumental events belong to one of the following groups:

- Wet End or Dry End Breakdowns (WEB/DEB)

- "Flashers" (light emitting PMTs)
- Electronics pickup events
- Flat TAC² events

These events can be identified with algorithms due to the specific characteristics of their event topology. The algorithms scan events and determine which events in a data file meet the criteria to be flagged as a specific characteristic. This process is known as data cleaning and subsequently the flags are known as data cleaning flags [44].

The 9 specific data cleaning flags used for this analysis follow below.

1. Fitterless Time Spread (FTScut): Cut designed to target "blind flashers"³ which typically occur when the readout from a specific PMT is disabled and the only leftover signal is from the produced light. To separate these from Cherenkov events we look for events which are relatively spread out in time since light emitted during a Cherenkov event would be \sim instantaneous. This cut starts with a list of all hit PMTs in the event with good calibration and compares the physical distance between PMTs in hit channels as well as their relative timing. Systematically pairs are created for channels which have timing $\delta t < 25$ ns and distance is < 3 m. If there are more than 15 valid pairs with a median $\delta t > 6.8$ ns the event is flagged.

2. Flasher Geometry Cut (Flashergeocut): Another pair-wise comparison, the flasher geometry cut uses spatial information from electronic space and detector space⁴ to find

²TAC - Time to Amplitude Converter: device that translates the time into charge (in ADC counts). It allows to calculate the time of a PMT hit just by measuring the stored charge. A Flat TAC event is one where the PMT hits are spread fairly evenly in time through the event.

³Blind flashers occur when the offending channel does not show a hit corresponding to the light emission

⁴detector space refers event information captured by the detector electronics pertaining to the geometry, hit timing and charge of the photomultiplier tubes relative to each other.

a burst of hits in close proximity to each other. For each hit, the ratio of all hits in a 1 m radius to the total number of hits in the event is compared and if it exceeds 1/2, the cluster is considered to be a possible flasher origin if the total number of hits in the region is also more than 4.

For each possible flasher origin the average hit time within the cluster is compared to the average hit time outside of the cluster. If the timing within the cluster is more than 500 ADC counts before the average hit time outside of the cluster or if one of the hits inside the cluster is a bad hit and the distance between the center of the cluster and the average position of the outer hits is ~ 12 meters (across the detector) the event is flagged. Another check is then performed in electronics space. The criteria here is that 4 hits must be within 8 channel numbers of a single hit. If a cluster is found the previous checks are performed again.

3. In Time Channel Time Spread (ITCtimespreadcut): This cut uses the hit timing profile of channels with good calibration. The percentage of hits in a sliding 93 ns time window is calculated and if the window with the highest percentage of event hits is less than 60% of the event hits the cut is applied. This targets "flat TAC" events.

4. Neck⁵: A simple cut, the neck cut targets events occurring in the neck of the detector. It uses the channels associated with the neck PMTs - if 2 neck PMTs are hit then the cut is applied. In the case where only one neck PMT is hit the spatial information is used to look for light in the bottom of the detector. If the average time between the

⁵During running, 3 PMTs were initially used in the neck before a channel was lost and only two were available. The Neck cut logic was unchanged

events in the bottom half of the detector and the neck hit is more than 85 ns after the neck hit, the event is flagged.

5. Charge Cluster (Qcluster): Targeting high charge events causing pickup in other channels, the qcluster cut uses electronics space data to see if more than 3/5 channels in a 5 channel wide window are hit. If any of these hits are considered bad⁶ then the event is flagged.
6. QvNhit: The charge of hits is compared to the number of hits in the event to remove events with very little charge in most of the hit channels. This should catch low level pickup events that are just above the trigger threshold. The charge from the lower 90% of hits is summed and converted to a total number of photoelectrons. If the total number of photoelectrons is less than 1/4 of the number of hits then the event is flagged.
7. QvT: Calibrated PMT data is used to identify flashers through a more relaxed means than the flashergeo cut since spatial information is ignored. The hit with the highest charge in two of the three charge channels (QLX and QHL) is considered to be a high charge hit if:

$$QHL_{\max} - \langle QHL \rangle \geq 610 \quad (5.1)$$

$$QLX_{\max} - \langle QLX \rangle \geq 110 \quad (5.2)$$

If the conditions are met, the timing of each of these channels is then compared to see

⁶a "bad" hit is one where the charge is un-physical in the charge integration channel. This is indicative of the wet-end breakdown, a flasher or a sharkfin (sharkfins occur when neighboring crate PMT channels experience electronic pickup due to a high charge hit in a nearby channel).

if it precedes the median by an amount that would indicate a flasher:

$$60 \text{ ns} \leq t_{\text{median}} - t_{\text{max}} \leq 250 \text{ ns} \quad (5.3)$$

Providing both conditions are met the event is flagged.

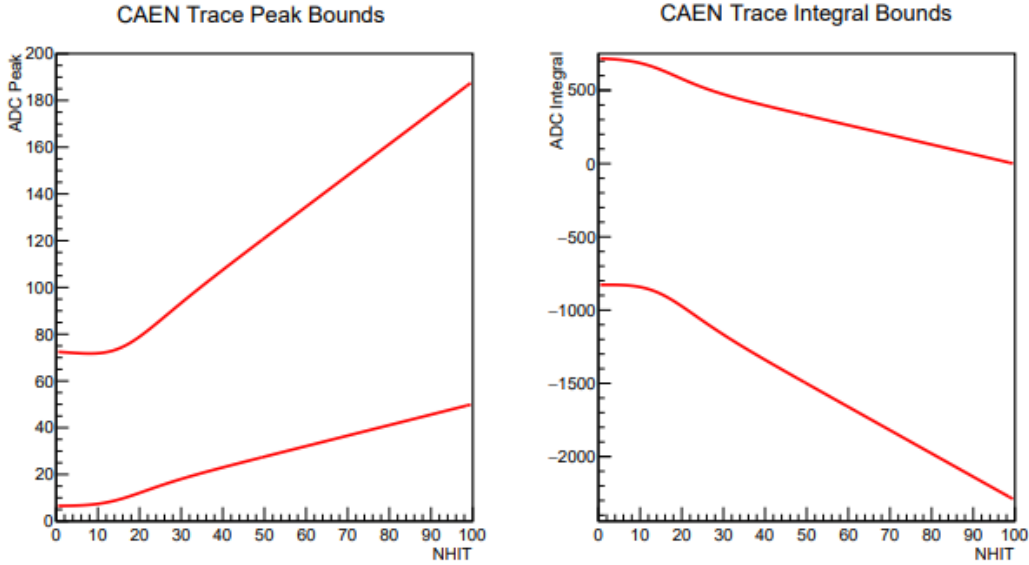
8. Ring of Fire (Ringoffire)⁷: The primary purpose of the Ring of Fire cut is to remove Ring of Fire events from the data. If 70% of event hits are in a single crate and 80% of those hits are in the outer ring of a crate the cut is applied. This cut is generally intended to identify electronic pickup on the periphery of the electronics crate.
9. CAENcut: CAEN digitizer data is used to compare the event ESUM profile to specific peak and integral bounds which are set with respect to the event Nhit. If these values fall outside of the bounds (see Figure 35) then they are flagged.

Since the burst files are written out with the same format as standard detector data, no modifications need to be done to the data cleaning processors in order to apply these data cleaning flags to events in burst files. For this thesis, a statistical analysis of data cleaning flags as they are applied to burst files is performed. Only some of the flags are interesting for burst files and are chosen for analysis; these flags have been described above. For all others an excerpt from the document "Status of Data Cleaning During the SNO+ Water Phase" can be found in Appendix A. Ultimately the goal of this analysis is to separate supernova bursts from bursts due to other causes.

⁷A ROF event is one where hit channels in crate space mostly occupy channels in the outer ring of a crate spatially. This cut is redundant with the crateisotropy and is intended to be removed in the future from the analysis mask

	T_{NHIT}	T_w	a	b	c
Peak (lower bound)	15	5	6.6	5.1	0.45
Peak (upper bound)	15	5	73.4	54.14	1.34
Integral (lower bound)	15	5	-834.0	-707.0	-15.9
Integral (upper bound)	15	5	720.6	658.8	-6.61

CAEN Cut constants for SNO+ water phase.



CAEN cut digitized peak (left) and integral (right) bounds as a function of NHIT.

Figure 35: CAEN Cut trace peak and integral bounds as they are applied to SNO+. If the ESUM profile is observed to be outside of these bounds the event in question is flagged by the data cleaning software.[44]

5.2 Data Handling

The main objective of this analysis is to observe trends and correlations within burst files that might be indicative of the nature of their cause. The cause of some bursts overlaps with the events that are removed from physics data with data cleaning flags. Since the data cleaning flags are designed to indicate the nature of the undesirable data it makes sense that they can be used in the analysis of bursts to indicate their cause.

After production, the burst files are stored on a local server until they are accessed for pro-

cessing/analysis. Next, a program reads in individual burst files and processes them through the SNO+ RAT data cleaning regime. Data cleaning is done in two separate passes since some cuts require comparison of event-event data to complete. This can consist for example of the first pass finding events which occur within a sliding time window of one another and the second pass flagging those events.

RAT is capable of modifying the data set to include or exclude parameters specific to the data set that it has processed. At its most basic level a RAT macro can read in a .zdab⁸ data file and output its unmodified contents as a ROOT file. Usually however the user would wish to process the data in some way. Depending on the bitmask written to the .zdab, the user would specify that they would like information about the detector target (relative combinations of air, water or LAB and the location of the interface) or timing also be written to the ROOT file. For this analysis all of the relevant metadata was written to the output ROOT file along with the data cleaning flags associated with events in the file.

After the data cleaning flags have been written to the output ROOT file, they can again be read out along with the event to which they apply. The ROOT macro used for this analysis first updates a text file list containing the total number of events in the burst file for normalization purposes. It then reads out, event by event, a list of the data cleaning flags which apply to the event in the burst file and writes the data cleaning flag words to a text file.

A Python script parses the text file, summing each occurrence of the data cleaning flags into a single component of a larger vector spanning all of the data cleaning flags. The components corresponding to the specific flags of interest are selected and compiled into a single

⁸.zdab was the file format associated with the SNO data structure. It will continue to be used by SNO+

vector representing the burst file from which it was generated.

This process is run continuously until each of the burst files to be processed has been processed and a list of vectors has been formed corresponding to the list of bursts which were to be processed. The vectors are stored in a .txt file from which the statistical analysis can begin.

Principal Component Analysis (PCA), which will be explained in detail in section 5.3, can be used to make a multi-dimensional representation of the data in our N-Dimensional space. Since the variance in magnitude of the burst vectors (\vec{V}) is quite large, it is useful to normalize them before performing the PCA. This can be achieved in more than one way but the method chosen for this analysis was to simply normalize to the number of events in the burst files. That is,

$$\vec{N} = \frac{\vec{V}}{E} \quad (5.4)$$

where \vec{N} is the normalized vector and E is the total number of events in the burst file. Like this, each component of the normalized vector represents the fraction of times a certain flag occurs per event in the burst file. All components are less than 1. This accounts for any large differences in magnitude between different bursts vectors which can occur due to greatly varying burst lengths.

ROOT comes with algorithms to compute principal components and was used to do so for the data set. The ROOT PCA package calculates the parameters for a transform to convert the original vectors components to the principal component space which can be used for better visualization of the data. The transform is computed and then the data transformed to the

new coordinate system before it is used to make plots.

5.3 Representation of Burst Data in Principal Components

A problem arises when one attempts to visualise a data set with dimensionality greater than 3. That is, as the number of dimensions increases it becomes impossible to create a representation of the data in two or three dimensions without losing certain specific components of the data itself.

Principal Component Analysis is a method by which high dimensional data can be visualized in a familiar looking representation without losing much interesting information. The PCA algorithm uses the existing data to create a new orthonormal basis optimized to maximise the variance in the data on lower ordered basis vectors. If any particular basis component of the original data set has variance, it would be represented in the generated principal components basis since the principal components will exist as some linear combination of the original basis components.

The process for generating the principal components involves a few different steps. Since the entire transformation consists of a simple rotation and translation of the coordinate system, the first step of PCA is to find the new origin point for the principal components. To do so, the mean values of each vector component are used as the center of the data.

Next, the algorithm uses the data itself to find a line passing through the new origin which maximises the mean distance between the origin and the projection of each data point onto the line. If these projections are summed in quadrature we simply need to maximize the sum. That is, it must maximize:

$$\sum_{i=1}^n \left| \frac{\vec{v} \cdot \vec{s}_i}{\vec{s}_i \cdot \vec{s}_i} \vec{s}_i \right|^2 \quad (5.5)$$

Once the line satisfying these conditions has been calculated, it is taken to be the first "principal component" of the data set. By definition it is the component spanning the greatest variance in the data and can therefore be considered to be the most informative component of the data. The other principal components are calculated in the same manner as the first principal component. Like this, a new orthogonal basis is calculated that spans the space of the original data set but whose axes are arranged from greatest to least variance span.

After the new basis, called pattern space, has been calculated one need only project the data set onto it to obtain the new PCA optimized data set. This is the general approach used for PCA however other methods exist which will obtain the same result. One could, for example, calculate the distances of the individual data points to the candidate axis and minimize the sum of these values. This is similar to the approach taken by the LINTRA based PCA package native to CERN ROOT installations [45] which was used for this analysis:

Each data point \vec{x}_n is read into the script as a column vector of N dimensions,

$$\vec{x}_n = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad (5.6)$$

where the subindex n refers to the specific data point in the set.

We then calculate a covariance matrix by comparing each component of each data point to the mean values of those components from the data set:

$$C_{i,j} = \frac{1}{n} \sum_{a=1}^n (x_{a,i} - \langle x_i \rangle)(x_{a,j} - \langle x_j \rangle) \quad (5.7)$$

When we assemble \vec{C} then we get an $N \times N$ matrix of the average variances for all of the vectors from which it was generated:

$$C = \begin{bmatrix} \langle C_{11} \rangle & \langle C_{12} \rangle & \cdots & \langle C_{1N} \rangle \\ \langle C_{21} \rangle & \langle C_{22} \rangle & \cdots & \langle C_{2N} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle C_{N1} \rangle & \langle C_{N2} \rangle & \cdots & \langle C_{NN} \rangle \end{bmatrix} \quad (5.8)$$

The eigenvalues and eigenvectors of the covariance matrix give the variance spanned by each principal component basis vector for the data set. A proof of this concept can be found in the ROOT TPrincipal class documentation [45].

ROOT's native LINTRA package calculates the transformation required for individual vectors based on the calculated eigenvectors, eigenvalues, mean values and standard deviations for the data set.

C++ uses one dimensional arrays for computation purposes so the entire set of eigenvectors is represented by a 1-D array where the 2-D array,

$$e_{i,j} = \begin{bmatrix} e_{0,0} & e_{0,1} & \cdots & e_{0,(N-1)} \\ e_{1,0} & e_{1,1} & \cdots & e_{1,(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ e_{(N-1),0} & e_{(N-1),1} & \cdots & e_{(N-1),(N-1)} \end{bmatrix}$$

becomes,

$$\vec{e}_s = \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_{N^2-1} \end{bmatrix}$$

using,

$$e_{i,j}^{\vec{}} \longrightarrow \vec{e}_s = \vec{e}_{N \cdot i + j}$$

The transformation obtained is:

$$T_j = \sum_{i=0}^{N-1} \frac{(\vec{x}_{n,i} - \langle \vec{x}_i \rangle) \cdot \vec{e}_{(i \cdot N + j)}}{\sigma_j} \quad (5.9)$$

Where j is the indexed number associated with the vector element being calculated. Working in this space requires that vector elements are indexed from 0.

The transformed data is then stored in an array where each vector has now been trans-

formed onto the axes of its principal components or in other words has been transformed into pattern space.

After this transformation has been performed, the basis of the pattern space will have components which are some linear combination of the original components. It is possible to select some number of these components to create a two dimensional representation of the data and retain much of the information from the original data.

The first calculated components of the set of principal components span the greatest variance in the data. It makes sense to choose principal components 0 and 1 to represent the data. The projection of all of the data points onto calculated principal components can be used in combination with each other to create scatter plots representing the data set. A scatter plot of the projections of the data onto principal component 0 and principal component 1 should reveal correlations and patterns in the data. For this analysis we expect that the PCA plots should show groupings within the data which would correspond to different causes of similar bursts.

6 Results and Interpretation

Principal component analysis was performed in several stages with a few different types of normalization to view the data in different ways. Since the sheer number of bursts generated between May 2017 and February 2018 (the sample studied here) would take a very several months to process, a representative sample spanning every 10th burst in 9 months was used for the analysis bringing the processing time down to just a few weeks. In addition, a subset of the data where only the first burst file generated in each sub run was analyzed in order to prevent many burst files that occur in the event of issues encountered by the detector from saturating the data.

Finally, the vectors contained in the 10th event sample were projected onto the principal components generated during SNO. Ultimately the work of Mike Schwendener[46] conducted during SNO motivated this study as it was useful in algorithmically determining the cause of bursts. This analysis was done in an attempt to reproduce a similar algorithm which could be used to once again algorithmically categorize bursts.

6.1 Data Flag Distributions

There is potential to gain some understanding of the burst mechanisms through the use of 1-D histograms. A 1D distribution of the data cleaning flags, both normalized and not, could give some indication of systematic phenomena causing different kinds of bursts.

1-D Histograms were generated for the distributions of data flags for a representative sample spanning every tenth burst file generated between May 2017 and February 2018 (Figures

36-44). We can see clearly that some of the flag distributions behave quite differently than others and may reveal more structure after PCA.

Each bin in the histograms represents the number of times that the associated flag was applied to an event in a burst file. The counts in a given bin show the number of burst files overall that had the specific flag applied as many times as the bin number in the histogram. Every distribution shows \sim singular peaks or otherwise regularly distributed bursts through all bins. No obvious anomalies or groupings can be seen in the data presented in this form.

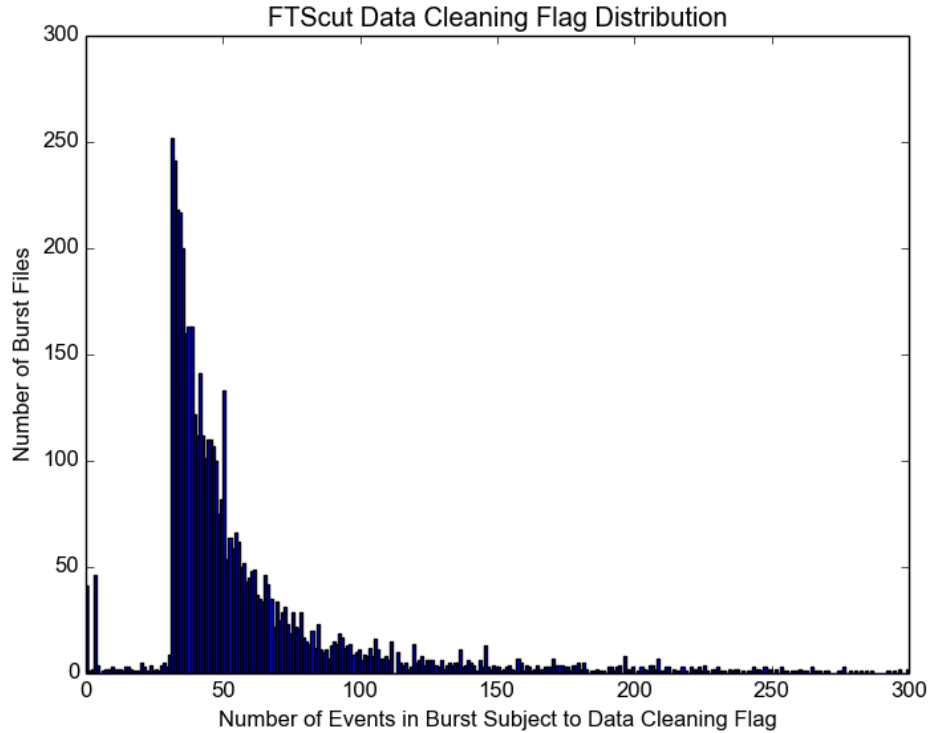


Figure 36: FTScut distribution - Peak at 34 flags per burst.

The FTS cut (Figure 36) shows a predominantly single peak with a few outliers that suggest some structure to the data. Of note here is the sharp rise at 34 events since this is the minimum number of events required to meet the burst condition by the L2 trigger. This

behavior suggests that the bulk of the information stored in the FTScut vector component is directly proportional to the number of events in a burst file.

Worth mentioning here however is that while the bulk of the information here is not indicative of the burst phenomenology, there are some interesting features outside of the main peak. The population of bins near zero and fifty suggest that there were bursts where the FTS cut parameter was not met. This would be indicative of certain bursts containing events of shorter duration rather than those spread out in time due to light signals travelling the diameter of the detector.

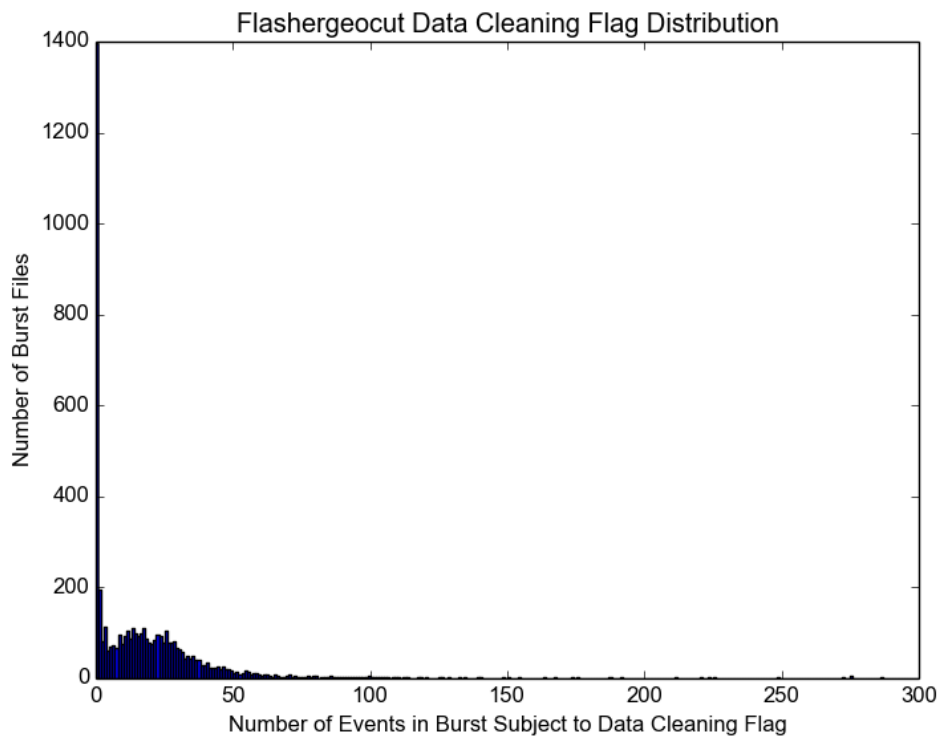


Figure 37: Flashergeocut distribution - Normal distribution with peak less than 34 flags per burst.

The bulk of the events in the flashergeocut (Figure 37) component of the data were in the

zeroth bin with a normal distribution centered around bin 20. The lack of a tail to the left is accounted for by the fact that this scale is absolute and so negative values are not possible, we would expect then that these events pile up in bin 0. Since the flashergeocut flags events where a cluster of events occurs 48.8 ns before the average hit which is nearly opposite the detector, this is indicative of this condition be a somewhat regular occurrence during bursts (and probably most detector running). The nature of this distribution describes a mostly uniform spread of flashergeocuts depending on degree rather than specific events. The few outliers could be indicative of bursts caused by repeated light emission from a PMT.

The itctimespread (Figure 38) cut uses a 93 ns time window to scan through the event in 1 ns increments discriminating against events where the hit distribution is "prolonged" while remaining somewhat uniform over time. The distribution is gaussian, centered about 23 events. This indicates that usually 60% or fewer of the events in a burst would have failed the itctimespread cut. This distribution is rather spread out however which indicates that as it would apply to a burst this varies by degree rather than according to the type of burst occurring.

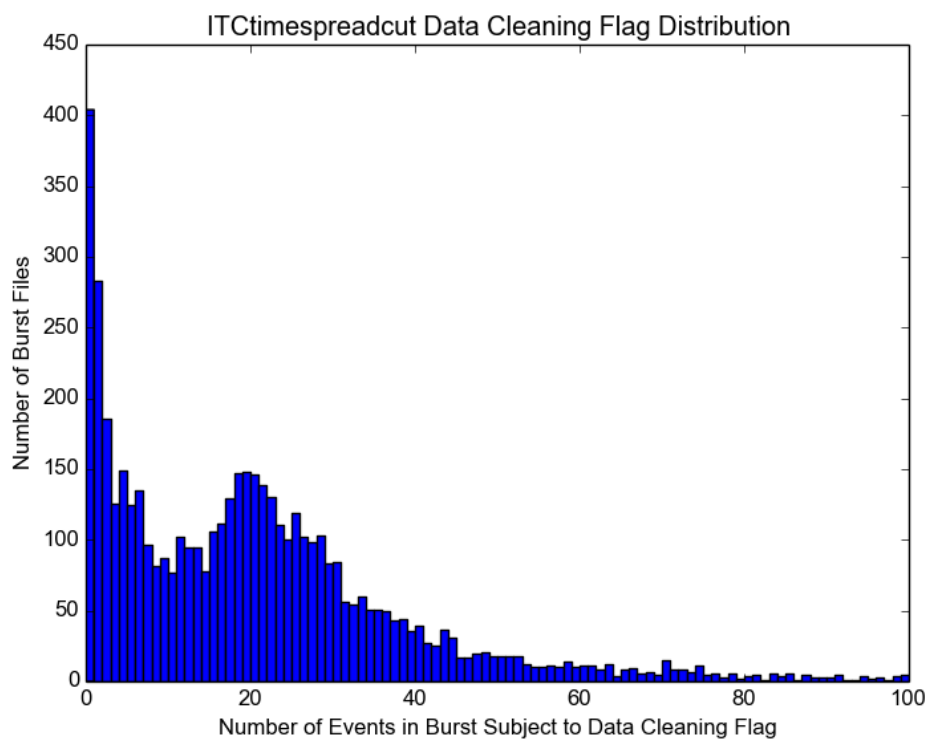


Figure 38: ITCtimespread cut distribution - Normal distribution with peak at 20 flags per burst.

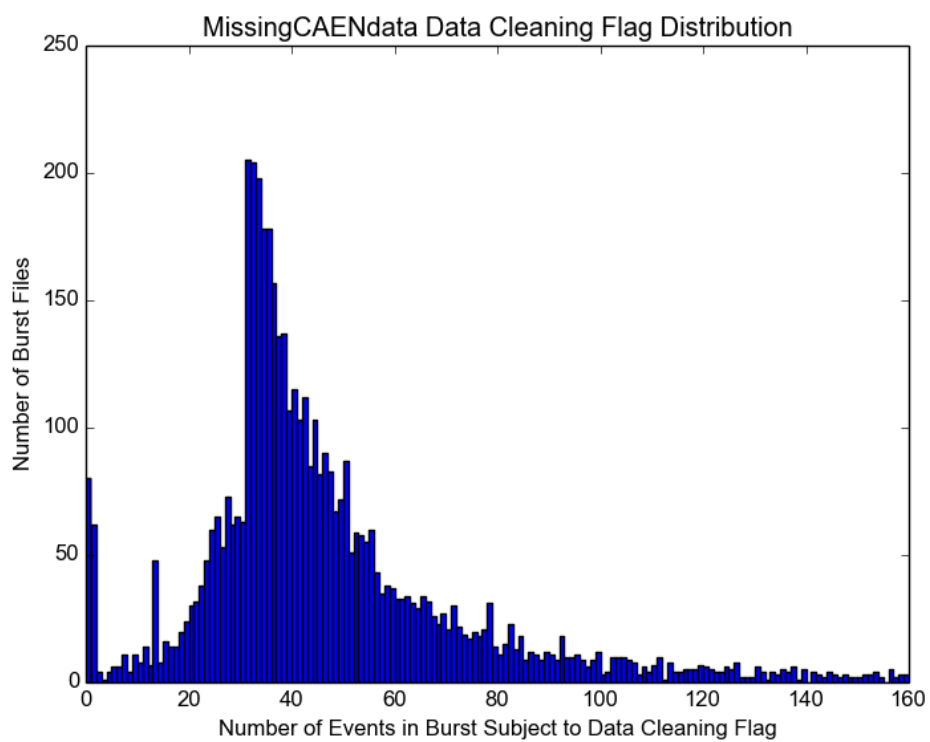


Figure 39: MissingCAENdata distribution - Bimodal distribution with peaks at 34 and 27 flags per burst.

The missingCAENdata (Figure 39) cut as its name suggests flags events where the CAEN data is not available or corrupted. This can occur due to occasional problems with hardware synchronization or with builder errors. During a high voltage breakdown⁹ event, it is not uncommon for the XL3 boards¹⁰ to lose timing synchronization. Such an occurrence could cause bursts which would have missing or corrupted CAEN data resulting in the flag being applied. The distribution has a prominent peak originating from ~ 34 events. Since this coincides with the minimum number of events required to create a burst candidate event it is reasonable to assume that there is a large portion of these events which are proportional to the number of events contained in the burst. Also, the significant number of data points falling around ~ 27 indicates that there is at least one other type of burst which doesn't simply saturate events with missingCAENdata flags.

The charge cluster (qcluster) cut uses electronics space information to target electronic pickup between channels due to high charge events in a single channel. This phenomena could occur due to electronic breakdown causing current through the electronics causing pickup in adjacent channels. Nearly all events in the representative sample fell in the zeroth bin (Figure 40). This indicates that this data cleaning flag bears no information for this analysis since the variance is zero. There are no outliers for this component.

The qvnhit cut uses electronics space information to determine the number of photoelectrons that would result in the charge that was observed from the lower 90% of hits in the event. If the ratio of the number of photoelectrons in the event to the number of channels in

⁹A high voltage breakdown is generally referring to the case where electronics contained within the PMT or in the dry end PMT power/signal distribution electronics begin to arc due to electrical component failure

¹⁰crate specific data readout electronics responsible for reporting crate status to the event builder when triggered by the trigger system over ethernet

that lower 90% is less than 0.25 (1 photon for every 4 channels) the event is flagged. This distribution from zero indicates a structured spread of the number of times the flag was called for each burst file (Figure 41). The exaggerated peak in the zeroth bin occurs due to a piling up of bursts which would represent a tail of the normal distribution.

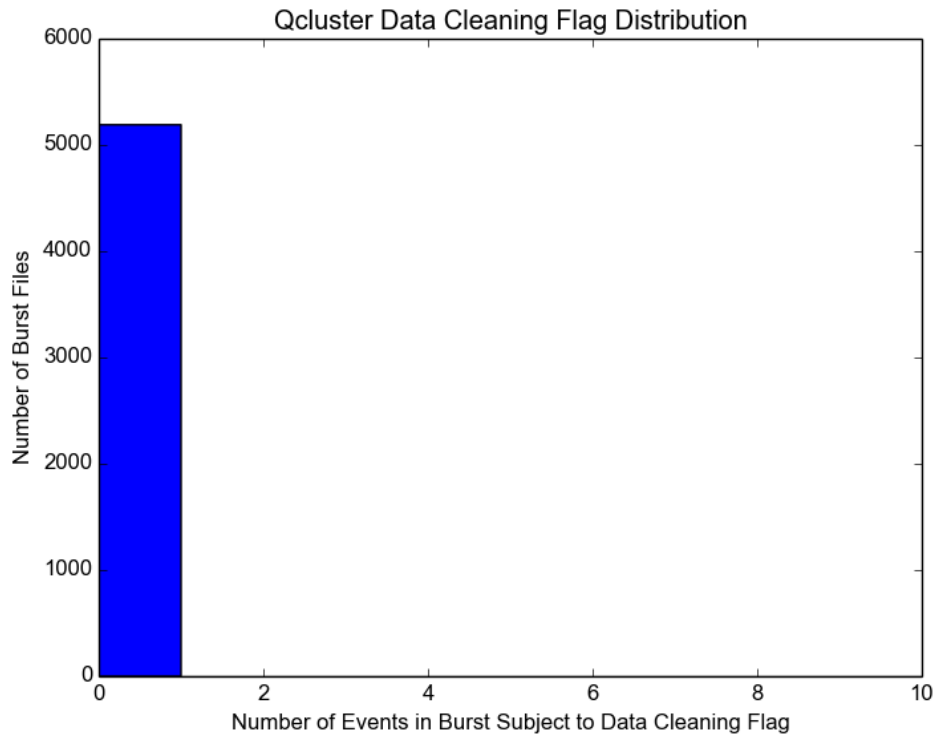


Figure 40: Qcluster distribution - all counts in zeroth bin.

The qvt cut compares the absolute time of the highest charge channel above a certain threshold which was hit in the event with the time of the median hit in the event. The sharp drop in the distribution (Figure 42) from zero indicates structure which may reveal more details when used in a PCA.

The ring of fire cut uses crate geometry information to flag events where the hits are concentrated in the outer ring of channels in a crate. This cut once again shows a quickly

fading distribution from zero which may reveal more details in PCA (Figure 43).

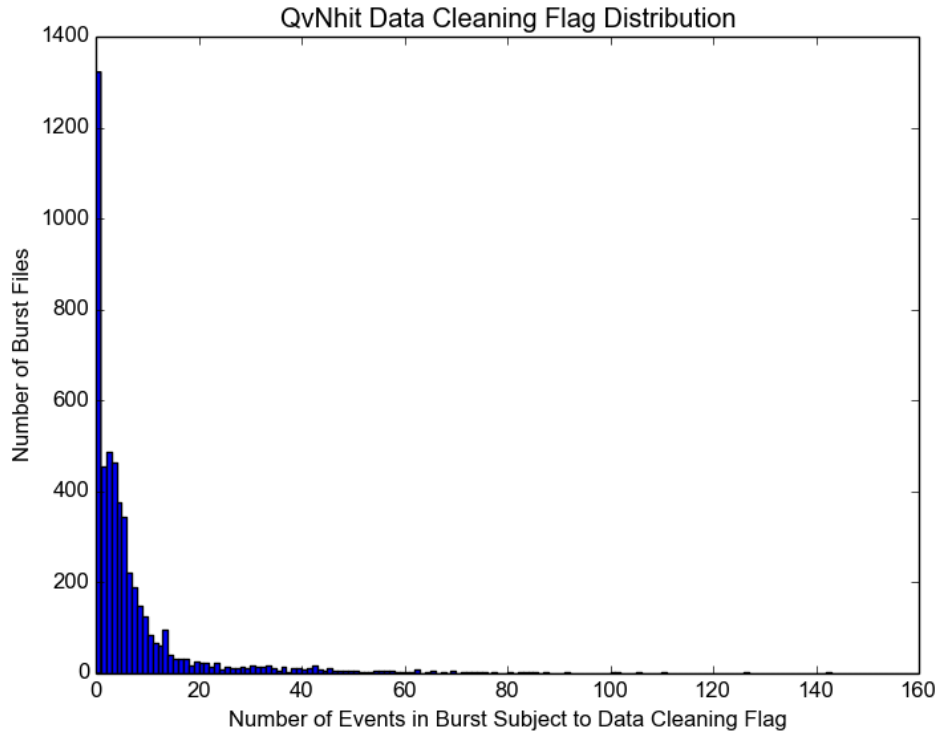


Figure 41: QvNhit distribution - Normal distribution with peak at 3 flags per burst.

The CAEN cut is applied based on the shape of the output pulse of the entire ESUM trigger signal of the detector for the event. This information may be vital for discriminating against different types of events which may occur and are not useful physics events. Once more we see a sharp drop in the distribution from the zeroth bin which may reveal more information when coupled to the other cuts through a PCA (Figure 44).

A distribution of the number of events contained in a burst file was created (Figure 45) to help explain some of the features in the data cleaning flag distribution plots. The counts of a specific bin represent the number of burst files containing a specific number of events. This histogram shows a peak at 32 events with another smaller peak at 16 events. The small peak

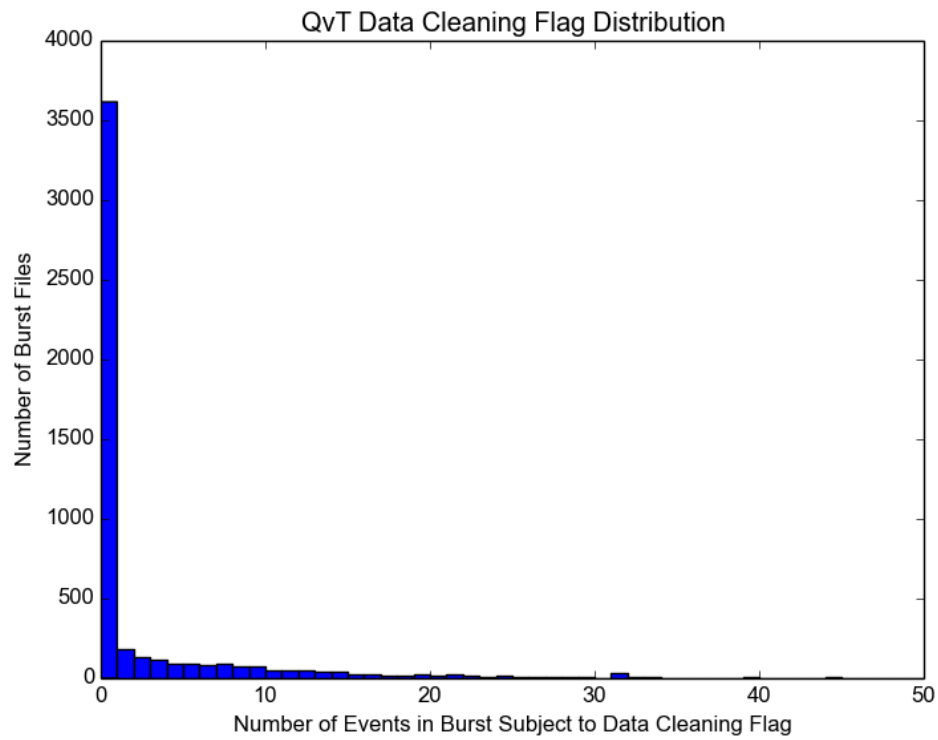


Figure 42: QvT distribution - Most counts in zeroth bin with a quickly fading tail.

at 16 events occurred all at once due to a small period of time where the L2 burst trigger was not set correctly. The minimum number of burst events in a burst file is typically 32.

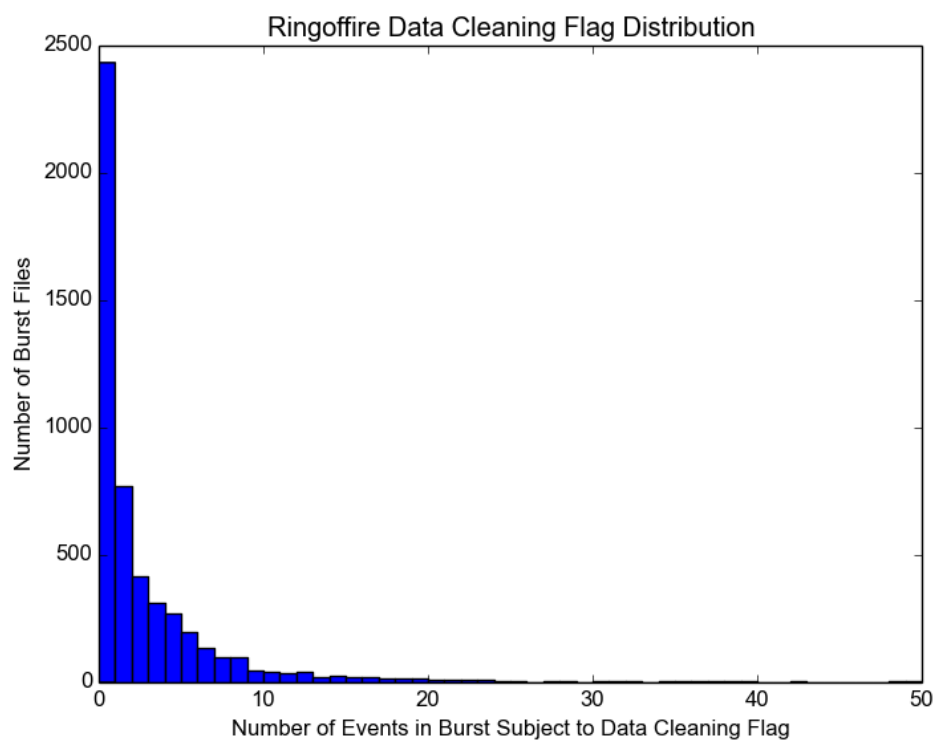


Figure 43: Ringoffire distribution - Most counts in zeroth bin with a prominent tail.

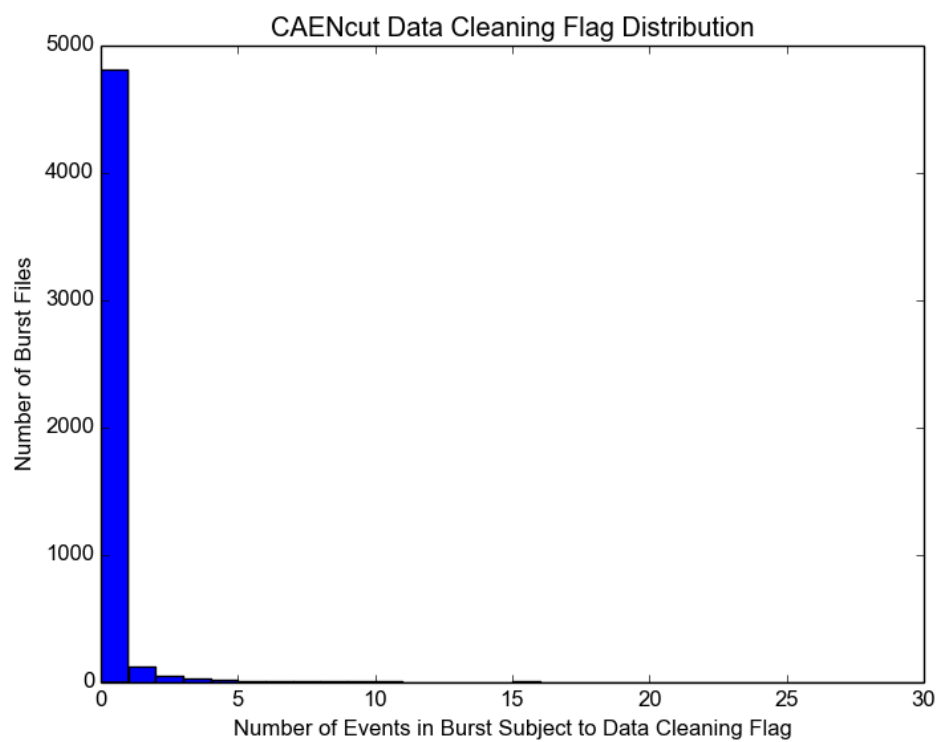


Figure 44: CAENcut distribution - Most counts in zeroth bin with a very weak tail.

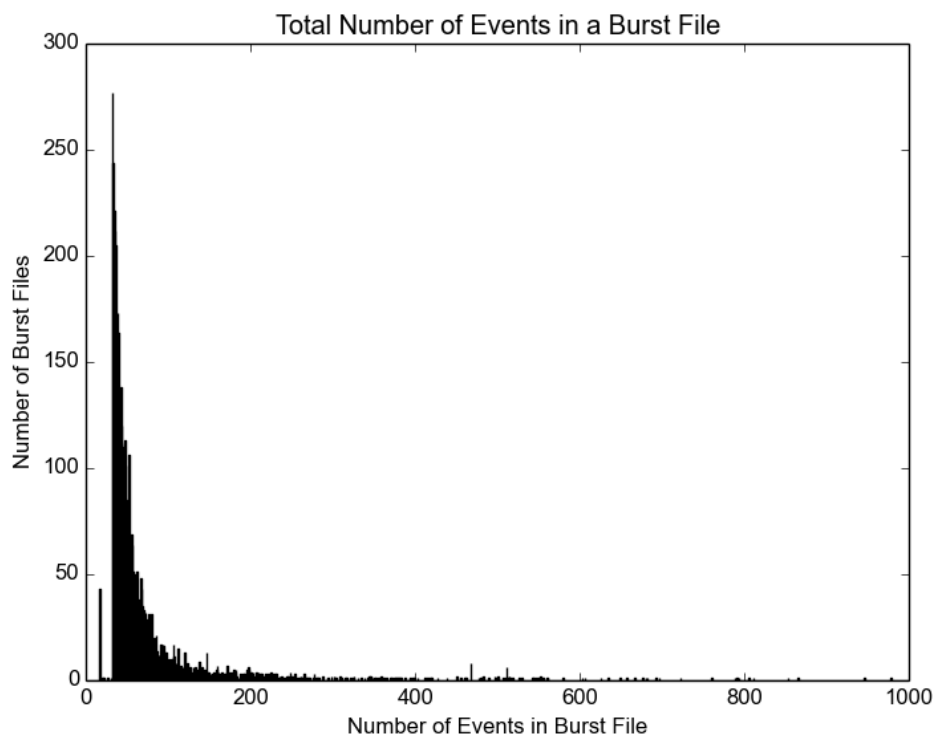


Figure 45: Distribution of the number of events in a burst file. The lowest peak occurs at 16 events and the main peak occurs at 32 events.

6.2 Principal Component Analysis

In the previous section we noticed that there was some structure within the groupings of individual data cleaning flag groups, but not enough to give any real indication that there were specific phenomena manifesting themselves in different ways through the detector. We can use a principal component analysis in order to re-map the data and tease out inter-dependencies within the data showing any structure between combinations of data cleaning flags for certain burst phenomena.

The principal component analysis was performed with a representative sample of the burst files spanning every tenth burst file. The representative sample was selected in order to decrease the processing time required to create the PCA plots without biasing the data selection.

The projections of each burst onto each principal component axis was plotted against the projection of the same burst onto another principal component axis. The most revealing plot was the projections of the data onto principal components 0 and 1 though the others can be referenced in Appendix C1.

Principal component analysis plots were also generated for other subsets of the data also normalized to the number of events in the specific burst. These can be found in appendix C2 Significant Bursts PCA Plots, C3 First Bursts PCA Plots and C4 All Bursts Projected Onto First Bursts PCA Plots.

As alluded to earlier PCA plots are the scatter plots created by projecting the data onto a set of principal component axes generated through calculation of the eigenvectors and asso-

ciated eigenvalues of the covariance matrix calculated from the data itself. The eigenvectors and eigenvalues corresponding with each principal component are listed below row-wise in Table 4.

Eigenvector									Eigenvalue
0.294	-0.477	-0.343	-0.120	0.103	-0.191	-0.186	-0.685	0.051	0.264
-0.093	0.455	0.114	0.601	0.118	0.149	0.289	-0.538	0.034	0.220
-0.362	0.153	-0.485	0.203	0.276	-0.529	0.266	0.375	0.033	0.149
-0.540	0.193	0.275	-0.231	-0.213	-0.105	0.009	-0.057	-0.695	0.128
-0.022	-0.290	0.495	0.531	0.007	-0.448	-0.418	0.110	0.005	0.100
0.409	0.264	0.087	-0.263	-0.358	-0.659	0.195	-0.271	0.112	0.057
-0.084	-0.561	0.248	0.072	-0.170	0.029	0.760	0.053	0.037	0.047
0.070	0.037	0.411	-0.356	0.814	-0.117	0.125	-0.077	0.013	0.030
0.550	-0.185	-0.274	0.217	0.185	-0.037	0.073	-0.042	-0.706	0.006

Table 4: Principal component vectors (eigenvectors of the covariance matrix) are displayed row-wise. The associated eigenvalue (fraction of total variance spanned by the eigenvector restated in Table 5) are to the right of each eigenvector.

Each row in the array above represents the basis components in data cleaning flag space of the principal component vector followed by the associated eigenvalue of that vector. The eigenvalue next to each principal component vector represents the fraction of the total variance of the data that the principal component spans. It makes sense then that the specific principal components used to represent the data show the maximum variance available in the data. The total variance that is shown in a 2 dimensional plot is given simply as the sum of the variance associated with each component used; with the same concept extending to 3 dimensions and so on. For ease of display we will choose to study only 2 dimensional representations of the data focusing just on the principal components which give the highest variance in the data. Tabulated we have:

principal component	variance fraction
0	0.264
1	0.220
2	0.149
3	0.128
4	0.100
5	0.057
6	0.047
7	0.030
8	0.006

Table 5: Variance of each calculated representative sample principal component

By default, the principal components will be arranged in order from highest to lowest variance making principal components zero and one the obvious choices for maximising the displayable variance in the data with eigenvalues of 0.264 and 0.220 respectively. This means that the two dimensional plot created by projecting onto principal components one and two will span collectively 48.4% of the total variance of the data. The combination which would show the next highest overall variance would be projections of principal component 0 and 2 spanning 41.3% of the total variance. For this reason, this section will only address the plot created by principal components 0 and 1 with the reader encouraged to review the other representations in appendix C.

The principal components chosen are in the form of unit vectors with the most dominant component relating directly to the corresponding parameter in the original data set. As such we can state the effectiveness of each parameter on the direction of the principal component since this is given by the magnitude of each component of the principal component. What we find is that principal component zero is dominated by the ringoffire cut and flashergeocut while component one is most dominated by missingcaendata, ringoffirecut and flashergeocut.

This means that the bulk of the variance of the data is contained in those three data cleaning flags. Since there is much overlap between the principal component vectors it is worth noting where they differ most. The components which differ the most are: *missingcaendata*, *itctimespreadcut* and *ftscut*. Since these components differ significantly from each other this is where we can expect most of the variance between the data projections to arise. It is expected that the most similar components would result in the most correlative behaviour of the plots where the most dissimilar components would give rise to features in the data.

Data Cleaning Flag	Principal Component 0	Principal Component 1	principal component 1 - 0
<i>ftscut</i>	0.294059	0.0929191	0.2011759
<i>flashergeocut</i>	0.477022	0.454836	0.022186
<i>itctimespreadcut</i>	0.34266	0.11385	0.22914
<i>missingcaendata</i>	0.120005	0.600996	0.480991
<i>qcluster</i>	0.10311	0.118055	0.014945
<i>qvnhit</i>	0.191385	0.148616	0.042769
<i>qvt</i>	0.185605	0.28889	0.103285
<i>ringoffire</i>	0.685429	0.53823	0.147199
<i>caencut</i>	0.0513828	0.0340606	0.0173222

Table 6: Relative weight of data cleaning flags for each vector component of principal component 0, 1 and 1-0

The orthogonal principal component 0 and principal component define a plane onto which we can project all of the data points. This is shown in Figure 46. It is worthwhile to note that all transformations from the old basis vectors to the basis of the principal component vectors are linear and do not skew the data in any way.

The principal component plot (Figure 46) shows some structure which would not be vis-

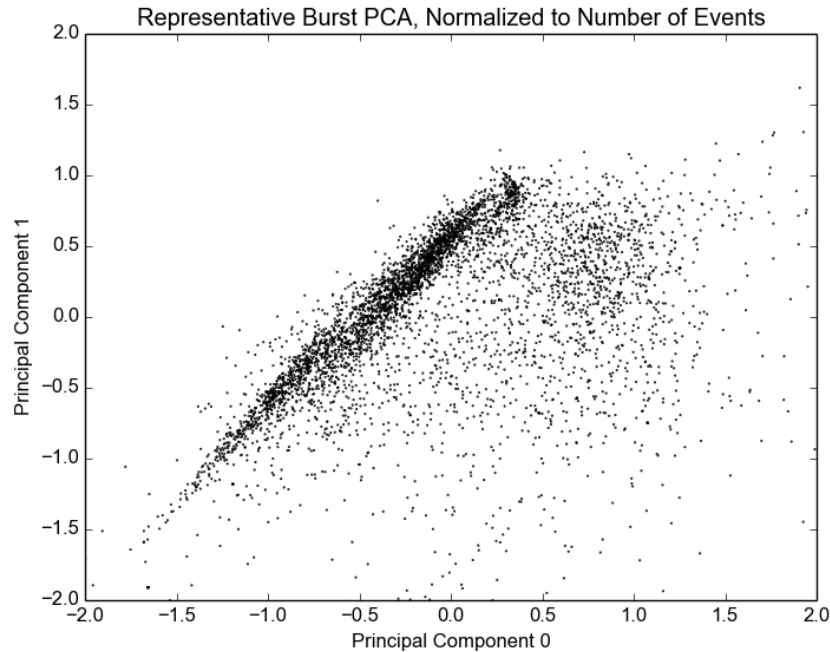


Figure 46: Principal component 0 vs 1, representative burst sample

ible while simply comparing the data cleaning flag information to each other. The plot separates the data in a few different regions worth mentioning. The first is the strong nearly one-to-one overall correlation between the two components with the lower region being populated by a more spread out distribution of data points. In this case it makes sense that there would be a strong linear correlation with this slope since the principal component vectors are dominated by a similar components. If there was a multitude of different arrangements of the data in the original vector form we would expect to see groupings in many different regions of this plot. We can gather then that in a most general sense the data is fairly uniform in nature which may be intuitive given the smooth distributions of data in the distributions of the original data cleaning flags.

Since the flashergeocut was expected to have a ratio proportional to the number of events

in a burst file and because the ringoffire cut is expected to follow the flashergeocut, it makes sense that this would in general represent an upper limit on this ratio explaining why most of the noise data points would fall below the prominent positive correlation region of the principal component analysis plot.

What is slightly more interesting than the general layout of the data in the two dimensional data is the noticeable three groupings laying within the strong positive correlative region (Figure 47). Unfortunately, the groupings are not distinct enough to postulate any sort of identification algorithm but they warrant further investigation nonetheless.

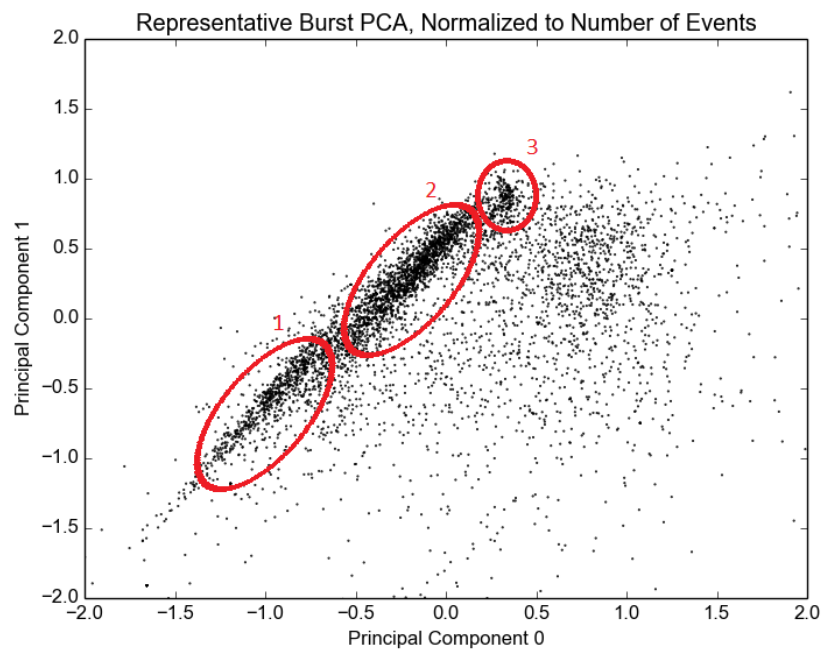


Figure 47: Principal component 0 vs 1 analysis representative burst sample. 3 distinct groupings are circled and labelled here

For this we will sift through the transformed vectors to find those where the two components are in the intervals $(-1.5 -0.5, -1.5 -0.25)$, $(-0.5 0.1, -0.25 0.6)$ and $(0.1 0.4, 0.6 1.1)$ for

each of groups 1, 2 and 3 respectively. A list of the names of the burst files from which these groups are composed can be found in Appendix B.

For now a list of the runs in which the burst occurred is created along with a note of the probable cause of the burst. This list is sorted into three columns based on the region in which the bursts can be found. The abbreviated list (Table 7/8) only considers runs where bursts were only produced in one of the 3 regions. In this form we can see that the First region generally will fill with events generated by detector running issues. More often than not these are shift in the baseline rate, detector ramping or synchronization issues of the detector electronics.

The second region shows some bursts which have been generated through slightly different means than the first region. Here we see more often detector electronics breakdowns and light introduced to the AV through the AV gate valves. This can be caused by simply opening the gate valves to allow a calibration source to pass into the detector or even light shone down into the detector through the UI¹¹ gate valves. In this form it would appear that region two contains light generating phenomena.

The third region contains very few bursts which were unique to the third region. From the table we might venture a guess that these bursts are arranged into the third region when a burst occurs that is in a detector running mode outside of the normal running mode for diagnostic purposes or ramping the detector.

Unfortunately, the lines between these regions are greatly blurred since this selective

¹¹The Universal Interface (UI) is a stainless steel cap sitting atop the neck of the acrylic vessel. It has a glove box and large actuated gate valves to allow workers to insert calibration sources and is sealed.

picture ignores phenomena which cause bursts in some combination of the three regions which make up the majority of the bursts which have occurred, it would be misleading to say that a burst which fits into one of the three identified regions of the PCA would have been caused by these associated conditions.

Run Number	Region 1	Region 2	Region 3
100102		Crate 12 breakdown	
100321	Detector ramping		
100550		Gate valves opened	
100956		Electrical Maintenance	
102008		Electrical work	
102036		Baseline shift	
102216	Timing Errors		
103315	Timing errors		
103670	Electrical diagnostics		
103688			Diagnostic running
103691	Diagnostic running		
103706	Detector ramping		
103740	Diagnostic running		
103762	Good run		
103764	Good run		
103836	Diagnostic running		
103981	Detector ramping		
104099	Crate 2 builder		
104632	Timing error		
104951	Shining light into AV		
104980	Diagnostic running		

Table 7: Probable cause of bursts during runs producing bursts in concentrated groupings. Continued in Table 8.

Run Number	Region 1	Region 2	Region 3
103315	Timing errors		
104632	Timing error		
104951	Shining light into AV		
104980	Diagnostic running		
103315	Timing errors		
105016	Timing errors		
105139	Baseline shift		
105262	Baseline shift		
105268	Baseline shift		
105288	Baseline shift		
105290	Baseline shift		
105725	Timing Error		
105812		Timing error	
105844		Timing error	
105846		Timing error	
105848		Timing error	
106672	Timing rack work		
106781	Timing error		
106782		LV/ramping	
107438	Baseline shift		
108242	Following Breakdown		
108531		Timing error	
108545		Diagnostic running	
108553		Wet breakdown	
108716		Rate spikes	
108948		HV Breakdown	
109080	Laserball transition		
109121		Swapping sources	
109126	Calibration induced		
109198		Normal operation	
109375		Timing rack reset	
109378		Low voltage	
109380			Detector Ramping
109418			Electronics work
109547		Electrical diagnostics	

Table 8: Probable cause of bursts during runs producing bursts in concentrated groupings. Continued from Table 7.

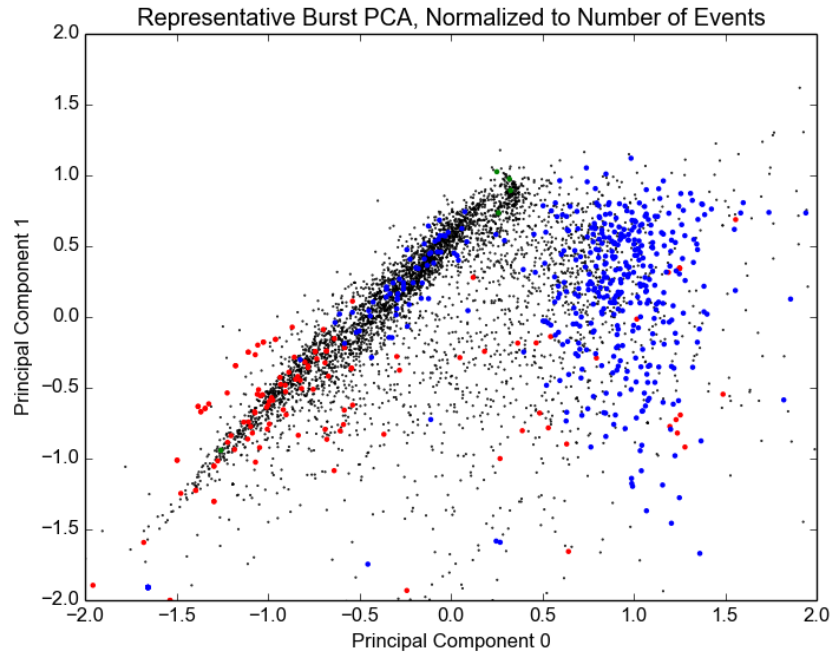


Figure 48: PCA with "unique"¹³ bursts highlighted. Bursts in run unique to group 1 in blue, group 2 in red and group 3 in green

When we highlight on the plot the bursts which occurred during runs with bursts falling only into one of the three regions ("unique" bursts), (Figure 48), we can see as many occurring outside of their respective regions of interest. Particularly, the case of the bursts associated with region 2 which are blue in the plot can be found outside of region 2. It is also worth noting that the outlying bursts of region 2 appear to make a second localised cluster in the plot. This suggests a binary-like grouping of these bursts.

Unfortunately the data is not closely enough correlative to draw any hard and fast conclusions about the nature of the bursts that have occurred in this period of water running after the detector had been re-commissioned. The data is very uniformly spread indicating that the bursts in general are flagged similarly enough to each other to be distinguished. This is

more or less confirmatory of the behaviour which was observed in the distributions of the application of the data cleaning flags in the previous section.

6.3 Bursts With Significant Content

The amount of physics data¹⁴ contained within a burst can vary greatly from burst to burst. Since the burst condition is met when the rate of production of events is high enough then the quality of data produced is determined by ratio of events passing all of the data cleaning cuts. We can then judge the extent to which a specific burst is significant by the number of total events in the burst which pass all of the data cleaning flags.

The data is already tabulated in the form of a ratio of the number of events in which a specific flag was called to the total number of bursts contained within the burst file. Because of this, we need only specify the degree to which we deem the burst to be significant. The data was generally overwhelmingly flooded with flagged events and so the tolerance must be set high enough to retain a reasonable fraction of the bursts which were used for the analysis. This ratio is tabulated in Table 9.

Though it is less than ideal, we will choose bursts in which $>2.5\%$ of all events pass the data cleaning cuts since this would retain 39% of the bursts sampled. The fact that such a low passing threshold was required in order to sufficiently populate a new PCA indicates that the bursts themselves do not contain much physics data. This confirms the idea that most of

¹⁴"Physics data" in this context describes events within a burst file which were not flagged by the data cleaning protocols.

Events Passing cuts (%)	Number of Bursts	% of all bursts
90%	2	0.04%
75%	4	0.08%
60%	15	0.28%
50%	21	0.40%
40%	29	0.55%
30%	44	0.84%
20%	67	1.28%
10%	143	2.75%
5%	186	3.59%
2.5%	2003	38.58%

Table 9: Percentage of bursts remaining, requiring a cleaning cut pass rate chosen from the representative sample of every 10^{th} burst file.

the bursts which were created were from short periods of time where the detector was not yet tuned perfectly.

The principal component vectors chosen for the PCA plot of significant bursts (Figure 49) are shown in Table 10.

Original PCs								
0.294052	-0.477022	-0.342666	-0.120005	0.10311	-0.191385	-0.185605	-0.685429	0.0513828
-0.0929191	0.454836	0.11385	0.600996	0.118055	0.148616	0.28889	-0.53823	0.0340606
Significant PCs								
-0.261832	-0.501725	0.308654	-0.154169	0.187702	-0.0420301	-0.427035	0.582102	-0.0498058
-0.222966	0.406683	-0.122328	0.592452	0.153711	0.17002	0.192177	0.573306	-0.0280226

Table 10: Original principal components vs significant physics data principal components

We can see that when comparing the principal components generated from the data containing significant bursts to those which were created using the entire data set that the principal components do not differ in magnitude from each other very much. This implies that the "significance" of the burst would not drastically alter the PCA and that the familiar shape of the new PCA plot (Figure 49) is reflective of the idea that the bursts are generally distributed

through the PCA as we see in the original PCA plot. Further plots featuring the data retained with a higher event data cleaning cut pass rate can be found in appendix C2.

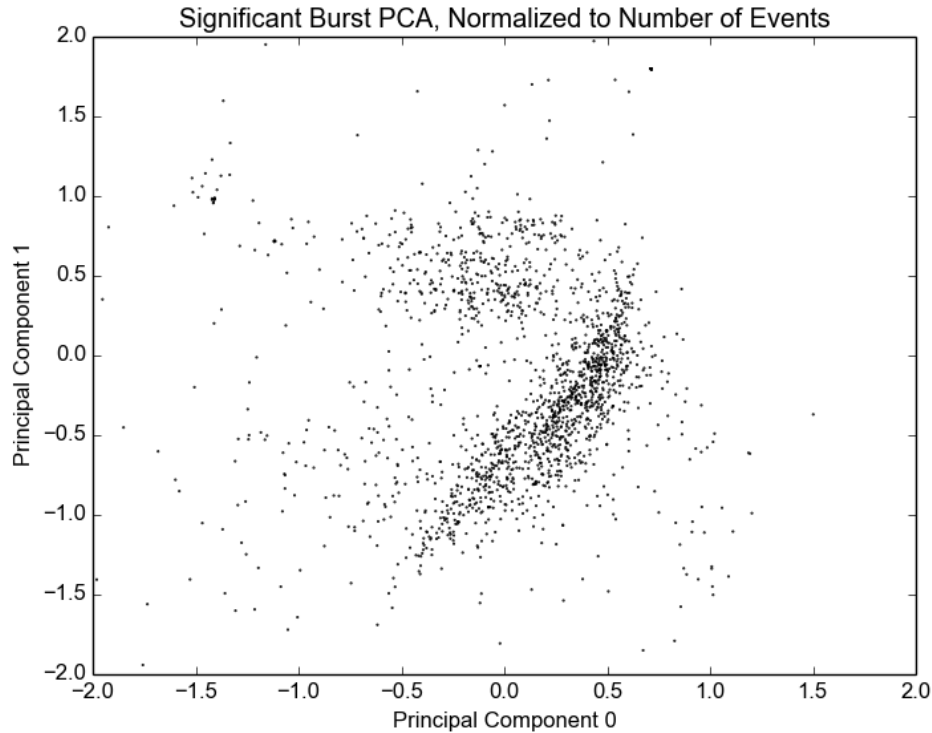


Figure 49: PCA with "significant" bursts. Bursts where more than 2.5% of all contained events pass all data cleaning cuts.

6.4 Removing Significant Backgrounds

Upon reviewing the comprehensive list of runs in which bursts are identified and the probable cause of the bursts, it becomes rather apparent that there is a significant background of bursts which were not caused by event-based phenomena but rather are due to the initial growing pains of detector tuning. Typically this can consist of a loss of timing synchronization between the trigger systems and the crate based data acquisition hardware, issues with the event

builder or other issues with the electronics systems. In retrospect this would be expected due to recent re-commissioning of SNO hardware, issues that arise from implementing new hardware and as the detector runs the operating crew would become more knowledgeable of how the detector behaves overall. This scenario prompted a third data set where only the burst file associated with the onset of a string of multiple consecutive bursts to be analysed.

The methodology used to remove these significant background events is to parse the list of burst files associated with a given run and eliminate bursts which occurred subsequent to the first burst within a single sub-run. Since we would not expect an interesting burst to be created in the midst of an extremely long and drawn out string of bursts we can remove these from the data set in a procedural fashion: The list of bursts is read into a script and only the first burst in a string of bursts from the same sub-run is used as a data point. This doesn't completely omit bursts due to hardware and software issues but it does remove long tails of bursts which would follow an event throwing off the detector state. As such we would be able to capture bursts which exist isolated from the rest of the data set as well as bursts which may have instigated the continuous spew of bursts from the detector. The burst files which make up this new list are intuitively called "first bursts".

The number of bursts analysed was reduced from 5180 to 685 first bursts corresponding to an 87% reduction in the number of bursts. For brevity's sake, only plots comparing the first four principal components of this data set were created since they would contain the majority of the variance in the data.

The first two principal component vectors chosen for the PCA plot of first bursts (Figure 50) are shown in Table 11.

Original PCs								
0.294052	-0.477022	-0.342666	-0.120005	0.10311	-0.191385	-0.185605	-0.685429	0.0513828
-0.0929191	0.454836	0.11385	0.600996	0.118055	0.148616	0.28889	-0.53823	0.0340606
First Burst PCs								
-0.515926	-0.00177008	-0.113383	0.3083	0.327919	0.185373	0.185839	-0.669938	-0.0257993
-0.183443	0.0426876	0.504532	-0.456285	-0.478545	0.0436451	0.456331	-0.248715	-0.027757

Table 11: Original principal components vs "First Burst" principal components. The first burst principal components are very different than the principal components calculated for the original representative sample bursts.

As we can see from the change in the principal components (Table 11), the "first burst" data cut changed the directions of the principal components significantly proving that the hardware issues encountered during this period would skew the results of the burst PCA and cause a background which could have been too great for any real phenomenological study to be conducted. It would seem then that the first burst data set appears to contain what we would expect to be more interesting bursts compared to the large original data set. Upon inspection of the PCA plot (Figure 50) however, we once again see a rather uniform distribution of the bursts without any significantly distinct groupings of like bursts.

Since the hardware issues could manifest themselves without a triggering phenomena, this data set is likely still heavily populated by a large fraction of bursts caused purely by these detector issues. It is likely then that more than 87% of the bursts analysed in this time period were caused by detector faults and that this method could be used to identify potential large strings of bursts such as these and systematically remove them from the data set. By projecting all of the data onto the first burst principal component axes, we can see that the data outside of the first burst group is separated from the first burst scatter plot (Figure 51).

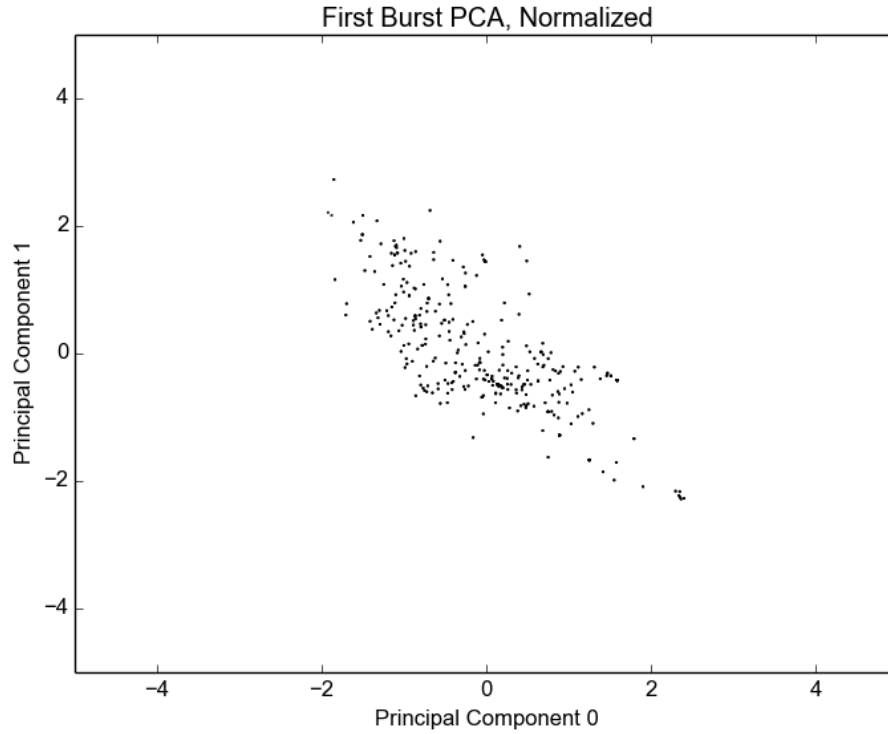


Figure 50: PCA with "first" bursts. Only the first bursts in a series of sequential bursts are analysed.

The most densely populated region on this plot does contain some overlap with the original data from which the principal components were created. Aside from this a linear tail can be found, indicative of the region where the bursts created as a side effect of poor running reside. It might be possible then to initiate a data cut based on the results of this projection with a marginally small loss of what could potentially be "real" bursts.

6.5 Comparison to SNO Data

A similar analysis was done using SNO data and similar data cleaning flags to those that were employed for this analysis [46]. The cuts which are essentially the same are the neckcut,

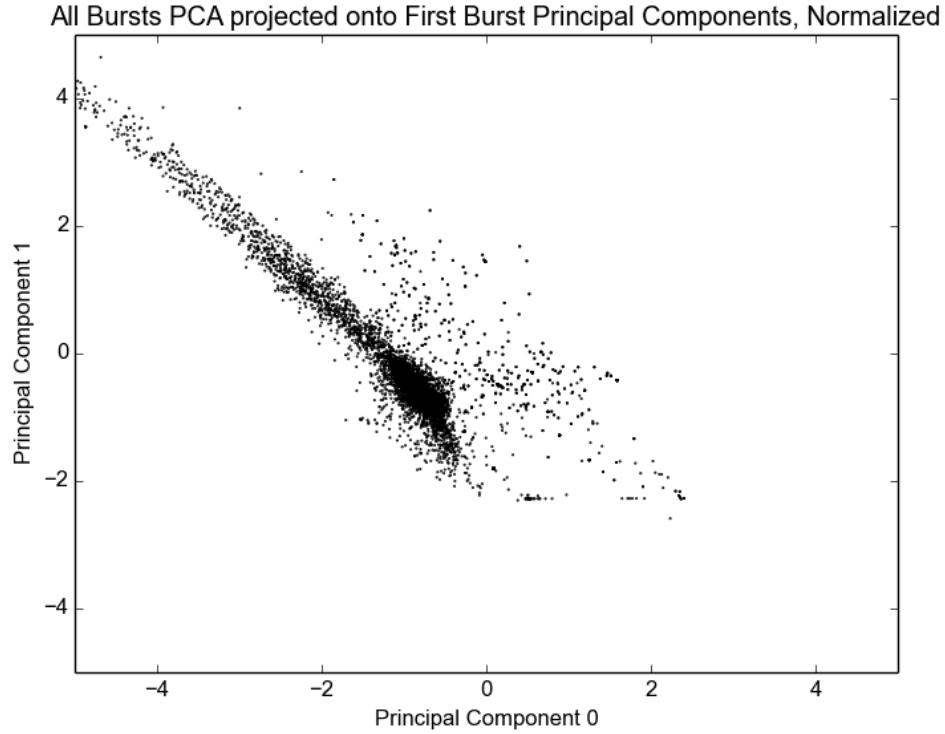


Figure 51: PCA with all bursts projected onto first burst principal components 0 and 1.

ringoffirecut, qvt, fitterlesstimespreadcut, qvnhit, crateisotropy, chargecluster, itctimespreadcut and the flasher geometry cut. It is worthwhile to note that cuts pertaining to the CAEN digitizer information are lost for this data since this was one of the upgrades to SNO+. This instead will be replaced by an instrument known as the AMB cut or analog measurement board which functioned similarly to the CAEN digitizer but passing less information to the DAQ. The first two principal components of this analysis are shown in Table 12.

We can attempt to transform our own data onto these principal components provided we make a minor change to these principal components in order to accommodate DAQ upgrades since SNO. We must note that this will not be principal components of this data set as they were not optimized from this data set. This is evidenced by the presence of nonzero values

Data Cleaning Flag	Principal Component 0	Principal Component 1
<i>ftscut</i>	−0.10	0.63
<i>flashergeocut</i>	−0.19	0.10
<i>itctimespreadcut</i>	−0.10	0.38
<i>ambcut</i>	−0.46	0.23
<i>qcluster</i>	−0.55	0.12
<i>qvnhit</i>	−0.03	0.18
<i>qvt</i>	−0.33	0.04
<i>ringoffire</i>	0.01	−0.15
<i>crateisotropy</i>	0.01	−0.07
<i>neckcut</i>	0.56	0.56
<i>eigenvalue</i>	0.36	0.23

Table 12: Burst data principal components during SNO [46]

for neckcut and crateisotropy cut which had to be eliminated from the analysis conducted for this thesis since neither contained any data other than zero values. Further we will assume that the AMB cut is similar enough to the caencut that they are interchangeable. Thus the vectors onto which we will project the data are shown in Table 13.

Data Cleaning Flag	Principal Component 0	Principal Component 1
<i>ftscut</i>	−0.10	0.63
<i>flashergeocut</i>	−0.19	0.10
<i>itctimespreadcut</i>	−0.10	0.38
<i>missingcaendata</i>	0.56	0.56
<i>qcluster</i>	−0.55	0.12
<i>qvnhit</i>	−0.03	0.18
<i>qvt</i>	−0.33	0.04
<i>ringoffire</i>	0.01	−0.15
<i>caencut</i>	−0.46	0.23

Table 13: Modification of SNO principal components to suit SNO+ burst data

Where the values taken to fill the missingcaendata slot have been filled with those from

the SNO neckcut in order to help preserve some of the orthogonality of these vectors. In this form the vectors will not be orthogonal but should be sufficient enough to give another perspective on the data. We will simply project the SNO+ burst data onto the SNO burst data principal component vectors to see if any patterns emerge (Figure 52).

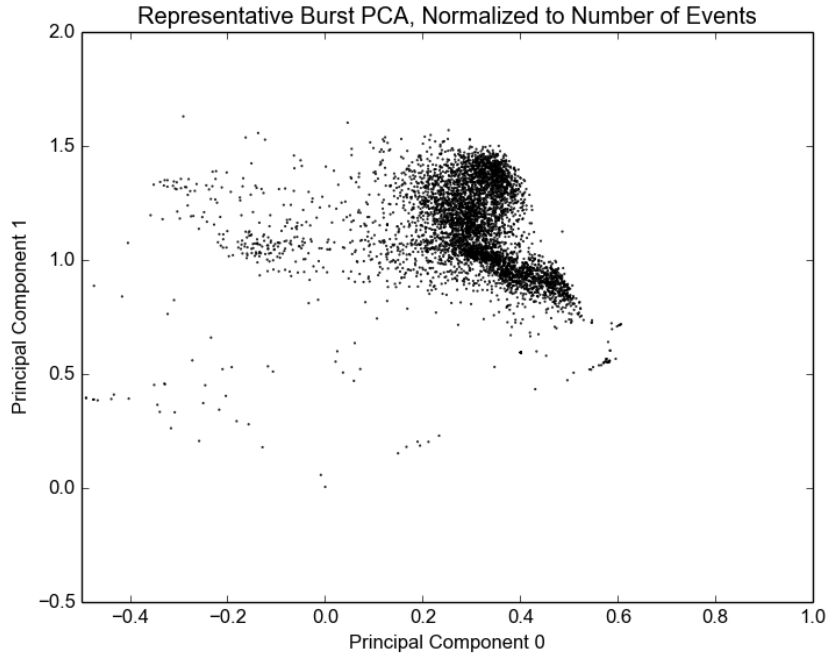


Figure 52: Principal component 0 vs 1 calculated during SNO. SNO+ burst data that has been normalized to the number of events in the burst file has been projected onto the SNO principal components.

This plot looks very similar to but not any more detailed than some of the higher order PC plots that were created from the original burst data. Most specifically it looks very reminiscent of the plot of principal component 5 vs principal component 6. In this way it is fair to say that this plot does not add any more detail to the original data than what could be seen with the optimized set of principal components. Also these basis vectors are no longer unit vectors or orthogonal to each other making this representation distorted compared to the optimized

basis vectors.

From the overall plot the reader would be forgiven for believing that there is some structure contained within the most densely populated region of the plot. Upon closer inspection of this region however (Figure 53) it becomes clear that any perceived structure would require further processing to tease out stronger patterns within the data.

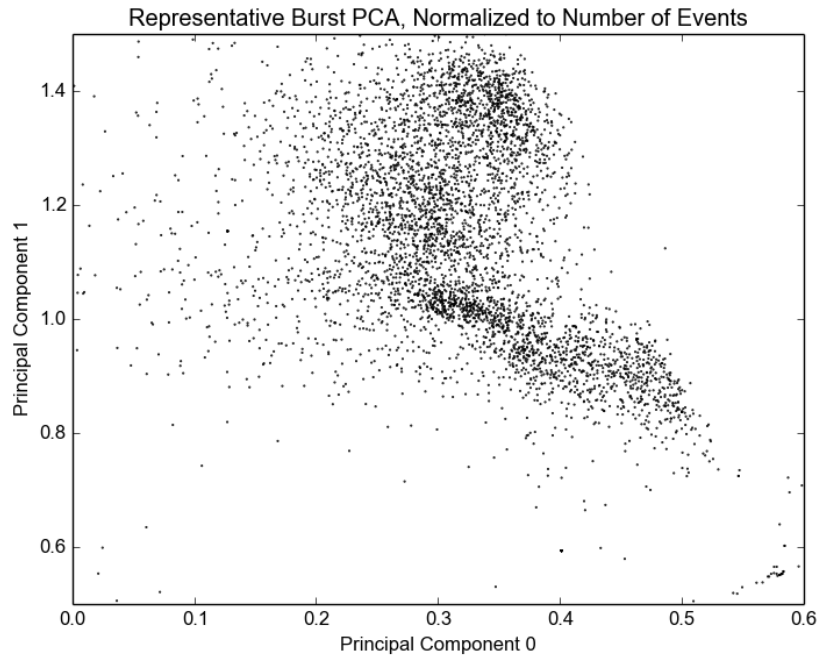


Figure 53: Principal component 0 vs 1 using SNO principal components. Normalized to the number of events in the burst file and showing the most densely populated region.

This distribution differs significantly from the SNO plot (Figure 54). This is most likely caused by a number of different changes made while upgrading from SNO to SNO+ such as the burst software trigger, the XL3 data readout cards, the presence of a light water target medium rather than the heavy water used during SNO and shifting baselines rates. This resulted in a higher than expected rate of production of bursts which in turn were not discrim-

inated by any data selection in the burst trigger. This meant that many bursts were created that looked very similar to each other preventing the distinct groupings that were seen during SNO being reproduced here.

At the time that this analysis was conducted, the data set was limited to the early data from detector commissioning and initial running. Over time the detector has been tuned and is now running much more smoothly than before. Repeating this analysis should yield different results which may in turn be able to algorithmically pinpoint burst pathology as it occurs. The analysis conducted here can be used to expedite future iterations of this analysis with more stable detector data and different detector target media.

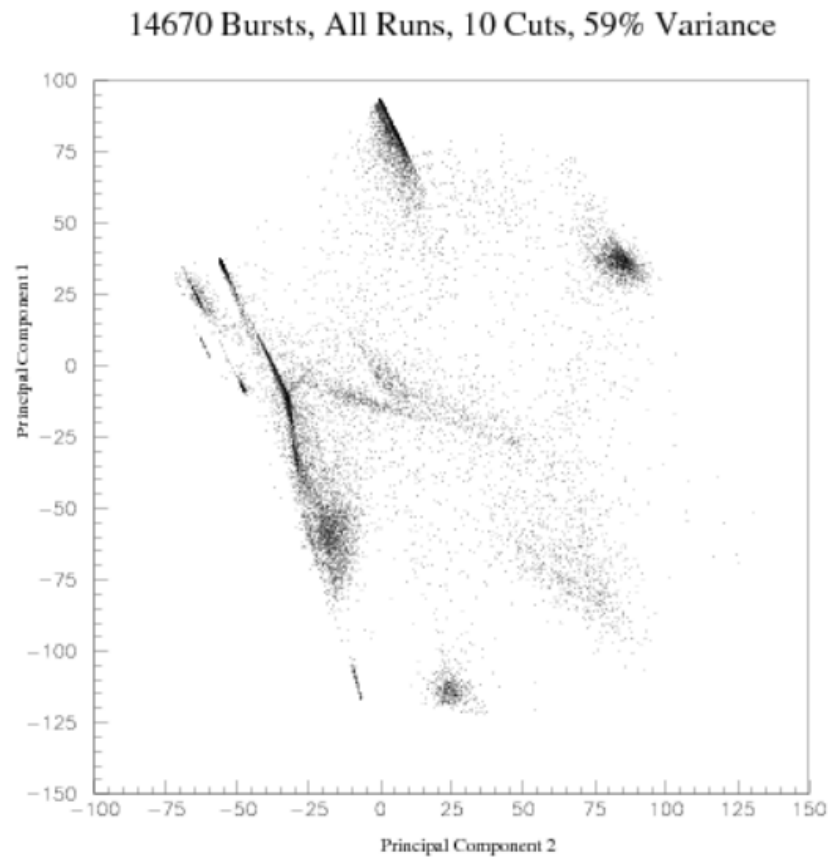


Figure 54: Principal component analysis of SNO burst data. [46]

7 Conclusion

Experiments in modern particle astrophysics typically require large and precise apparatuses. A large community of collaborations exist in the field of neutrino physics with the intention of further exploring the properties and processes associated with neutrinos. Weakly interacting solar neutrinos were detected by the SNO experiment in an attempt to explain the notable discrepancy between a predicted neutrino flux and a detected neutrino flux. The detection and analysis showed that neutrinos oscillated between flavour states and had mass earning a Nobel prize for the discovery [47].

Another astronomical phenomena for which the neutrino could be exploited because of its weakly interacting nature is the supernova. Stars progressing to the end of their life begin a chain reaction of nuclear burning until the thermonuclear balance of gravitational pressure and electron degeneracy pressure is upset resulting in a notable end to the star's life as a supernova. Supernovae are expected to emit large bursts of neutrinos which, if sufficiently close to the earth, would cause large signals to be produced by neutrino detectors worldwide.

SNO+ is one such neutrino detector and exists as a re-purposed version of the SNO detector. The kilo-tonne scale underground neutrino detector will go through several phases researching different specific neutrino physics goals with a primary interest in neutrino-less double beta decay in ^{130}Te . During all phases the detector would be capable of detection of a supernova neutrino burst and so provisions have been made to deal with the associated signal. Algorithms are currently in place which compile data known as "bursts" into "burst files" in parallel with the primary data compilation. The burst files consist of a minimum

number of detected events occurring within sufficiently small temporal spacing. This data could be instrumental in revealing more about the nature of the supernova phenomena. It is exceedingly important that this data be retained and studied if it has been collected.

Over the period of interest spanning May 2017 to February 2018 more than 52000 burst files were generated for reasons which were not related to supernova events. These bursts constitute a background for supernova data which may be useful to understand in order to expedite the process of identification for burst diagnostics. The data cleaning flags which are normally used to flag undesirable events in regular physics running were used to create vectors associated with individual burst files. Principal component analysis was performed on these vectors to represent the data, aiding in pattern recognition within the data set.

The distributions of the data flags were analyzed in both their raw and normalized form identifying that each flag usually consisted of either a single peak without any odd features or secondary peaks. This suggests that data cleaning flags were distributed rather uniformly amongst the bursts files.

Principal component analysis was then carried out on various configurations of this data set in an attempt to pull out any less obvious patterns which may be present within the data. The most basic configuration was of a representative sample of the data set spanning every 10th burst file, which appears to contain three main features associated with detector tuning issues, electronic breakdown and light introduction, or bursts during alternative running modes. There was however significant overlap between these regions meaning that this categorization would not be sufficient to label a burst of unknown origin. Bursts occurring during electronic breakdown and light introduction events however appeared to split into a fourth

grouping without any discernible cause which may prompt further study.

A second configuration of "significant" bursts defined the burst based on their ratio of events which were flagged to events which were not flagged. In order to retain a statistically significant fraction of the bursts the requirement was that a minimum of 2.5% of all of the events in a burst file must pass all of the data cleaning cuts. This resulted in 38.58% retention of all of the bursts from the original data set. The optimized principal components were not significantly changed from the original PCA and the distribution was strikingly similar to the original data as well.

A significant background was noticed in that a burst generating event could initiate poor detector behaviour for a sustained duration during which many burst files would be generated. In order to remove these, only the first burst would be retained in a new PCA. This resulted in significantly changed principal component axes and a 13% data retention. This implies that the bursts generated due to improper detector behaviour constitute an extremely significant background to identification of phenomena. Further, all bursts were then projected onto the axes optimized to the first burst group and found to congregate in a linear fashion near to the main cluster produced by the background reduced PCA. This suggests that this specific transformation could be used to isolate and remove bursts due to poor detector running at the cost of losing a small amount of the good data.

Finally a representation of the principal components used for a similar analysis during SNO were created reflecting the hardware changes that had occurred to upgrade from SNO to SNO+. Projection of the new bursts onto the SNO principal components yields a scatter plot forming a similar cluster to what can be seen in the original PCA plot. Since the

SNO data organized itself into many distinct groupings, we can conclude that there are fundamental differences between the data collection method used for the SNO PCA. Changes to the process systems, target medium, control electronics and data processing caused bursts in the SNO+ detector to manifest themselves with a higher rate and with a more uniform phenomenology than they had in the SNO detector.

This study shows that the method of principal component analysis can be applied to bursts produced in the SNO+ detector and reveal structure which may not be otherwise apparent. As time progresses and the detector is run for longer it can be expected that these bursts will become more distinct and reflective of the specific phenomena rather than detector issues. Further, the change of target medium from ultra pure water to the liquid scintillator linear alkyl-benzene should serve to further change the distributions of these flags throughout the generated bursts. This analysis should be repeated after the detector has been running for a sustained period with a scintillator target to further explore the phenomena which would cause bursts. If the segregation of bursts from different phenomena is sufficiently significant, this could then be used in parallel with detector operation to offer on-line burst diagnosis for detector operators.

References

- [1] MissMJ, *Standard Model of Elementary Particles*. Digital image, Wikimedia Commons, 2006.
- [2] W.S.C. Williams, *Nuclear and Particle Physics*. Oxford Science Publications, 2018.
- [3] Harp, *Quark structure of a neutron*. Digital image, Wikimedia Commons, 2006.
- [4] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*. Physics Letters B, 716(1):1 – 29, 2012.
- [5] Viktor T. Toth, *On how an up quark can be changed with a W boson*. Digital image, Quora, 2017.
- [6] W. Pauli, *Dear radioactive ladies and gentlemen*. Gauverein meeting in Tübingen, 1930.
- [7] J. Boger et al, *The Sudbury Neutrino Observatory*. Nuclear Instruments and Methods in Physics Research A, 449:172–207, 2000.
- [8] B. Aharmim et al, *Low Multiplicity Burst Search at the Sudbury Neutrino Observatory*. arXiv:1011.5436, 2010.
- [9] The SNO Collaboration, *Construction of the SNO Detector*. Sudbury Neutrino Observatory, Sudbury, Ontario.

- [10] C. Darrach, *The SNO+ Supernova Calibration Source Development and Testing*. MSc. Thesis, Laurentian University, 2016.
- [11] A. Pocar et al, *Solar Neutrino Measurements*. arXiv: 1812.02326, 2018.
- [12] Dorottya Szam, *Proton Proton Cycle*. Digital image, Wikimedia Commons, 2012.
- [13] Sonia Bacca, *Topics in Nuclear Astrophysics*. TRIUMF Graduate Course Material, PHYS 505, 2016.
- [14] S. Andringa et al., *Current Status and Future Prospects of the SNO+ Experiment*. Advances in High Energy Physics, 6194250, arXiv:1508.05759, 2015.
- [15] Shantanu Basu, *Astronomy 021 Lecture Slides*. Digital image, University of Western Ontario, Undergraduate Course Material, 2007.
- [16] European Southern Observatory, *Hertzsprung-Russell diagram*. Digital image, ESO USA, 2007.
- [17] M. Seeds, D. Backman, *Foundations of Astronomy 14th Ed*. Brooks/Cole Publishing, 2019.
- [18] Chaisson, McMillan, *Astronomy*. Digital image, Pearson Prentice Hall, 2005.
- [19] Chemistry, *Converting Mass to Energy: Mass Defect and Nuclear Binding Energy*. Libretexts, 2018.
- [20] The Hubble Key Project Team, The High-Z Supernova Search Team, *SN-1994D*. Digital image, NASA/ESA, 1994.

- [21] M. Turatto, *Classification of Supernovae*. arXiv:astro-ph/0301107, 2003.
- [22] R. Buras, M. Rampp, H.-Th. Janka, and K. Kifonidis, *Two-dimensional hydrodynamic core-collapse supernova simulations with spectral neutrino transport*. Astron. Astrophysics, 2006.
- [23] Department of Physics and Astronomy, *Particle Astrophysics*. Digital image, University of Alabama, 2018.
- [24] K. Hirata et al, *Observation of a Neutrino Burst from the Supernova SN 1987a*. Physical Review Letters, vol. 58, no. 1490, 1987.
- [25] L.V. Krivosheinas, *SN 1987A – historical view about registration of the neutrino signal with BAKSAN, Kamiokande II and IMB detectors*. International Journal of Modern Physics, 2004.
- [26] SNOLAB, *SNOLAB Technical Reference Manual*. Internal Reference Document, 2016.
- [27] Z. Petriw, *An Underwater Six-Camera Array for Monitoring and Position Measurements in SNO+*. Digital image, University of Alberta, 2012.
- [28] R. Ford, *A scintillator purification plant and fluid handling system for SNO+*. Digital image, American Institute of Physics Conference Proceedings, 2015.
- [29] Allison Currie Beaulieu, *Submission for SNOLAB photo walk*. Digital image, SNOLAB, 2018.

- [30] R. J. Boardman, *The photocathode area of the Hamamatsu R1408 photomultiplier*. Oxford nuclear physics laboratory, 1991.
- [31] J. Klein, M. Neubauer, F Mitch Newcomer, R. Van Berg, *The SNO trigger System*. The SNO Collaboration, 1997.
- [32] A. Mastbaum et al., *SNO+ Detector Operator Manual*. Internal document, The SNO+ Collaboration, 2018.
- [33] M. Anderson et al., *Search for invisible modes of nucleon decay in water with the SNO+ detector*. Phys. Rev. D 99, 032008, 2019.
- [34] M. Anderson et al., *Measurement of the ^8B Solar Neutrino Flux in SNO+ with Very Low Backgrounds*. Phys. Rev. D 99, 012012, 2019.
- [35] M. Redshaw et al., *Masses of ^{130}Te and ^{130}Xe and double—decay Q value of ^{130}Te* . Physical Review Letters, vol. 102, no. 21, 2009.
- [36] F.T. Avignone III et al., *Double Beta Decay, Majorana Neutrinos, and Neutrino Mass*. Rev.Mod.Phys.80:481-516, 2007.
- [37] R.N. Cahn et al., *White Paper: Measuring the Neutrino Mass Hierarchy*. Inspire HEP, arXiv:1307.5487, 2013.
- [38] S. Bilenky, *Introduction to the Physics of Massive and Mixed Neutrinos*. Springer, 2010.
- [39] M. Tanabashi et al. (Particle Data Group). Phys. Rev. D 98, 030001, 2018.

- [40] T. Kaptanolgu, R. Lane, *Spectrum Plot 0.5 Percent Loading using Te-Diol*. Internal document, The SNO+ Collaboration, 2019.
- [41] T. Kaptanolgu, R. Lane, *Pie chart for ROI contributions in year one for Te-Diol scenario*. Internal document, The SNO+ Collaboration, 2019.
- [42] K. Labe, *Report and User Guide for the Level II Trigger System*. Internal document, the SNO+ Collaboration, 2016.
- [43] The SNO+ Collaboration, *SNO+ Acronym list*. 2019.
- [44] C. Grant et. al, *Status of Data Cleaning During the SNO+ Water Phase*. Internal document, The SNO+ Collaboration, 2018.
- [45] C. Holm, *TPrincipal Class Reference - ROOT Reference Guide*. CERN, 2000.
- [46] M. Schwendener, *Supernova Monitoring in the SNO Detector*. MSc. Thesis, Laurentian University, 2002.
- [47] *Press release. NobelPrize.org. Nobel Media AB 2019. Thu. 15 Aug 2019.*
<https://www.nobelprize.org/prizes/physics/2015/press-release/>, 2015.

A SNO+ Data Cleaning Flag Descriptions

The following sections contain excerpts from the document Status of Data Cleaning During the SNO+ Water Phase[44] presented to the SNO+ collaboration by the SNO+ data cleaning group. It gives a full outline of the data cleaning protocols and how they affect the data which are retained for study by SNO+.

A.1 Zero-Zero

The zero-zero cut is the simplest data cleaning cut which simply removes any event in which the global trigger ID (in hex) ends in 00. This cut is put in place to avoid the effects of a rollover issue with the global trigger registers that will create orphan hits with bad global trigger IDs. Since 00 occurs at the end of sequential events every 256 events, there is an automatic (unbiased) sacrifice of 0.39% of the data from this cut alone.

A.2 Crate Isotropy

This cut targets events that are caused by pickup from the front-end electronics. This cut only looks at hits in electronic space (no consideration of location of PMT in detector, just where the crate/channel/card location). In order to fail this cut, an event has to first have most of its hits within a single crate. We define this as,

$$\frac{Hits_{in\,any\,one\,crate}}{TotalHits} > 0.7$$

Furthermore, if the event does pass this criteria, it must further exhibit signs of card-to-card or channel-to-channel pickup. Within the crate of interest, two comparisons are made. First

we check to see if any two side-by-side cards contains more than 80% of hits in that crate, and then we check if any two side-by-side channels (across all cards) contains more than 70% of the hits. If either of these two is true, then the event is removed from data.

A.3 Fitterless Time Spread

The fitterless time spread cut (FTS cut), is designed to look for so-called blind flashers. This is important when the suspect PMT has its readout disabled and thus we are unable to see the high-charge PMT that causes the flasher event and we are only left with the produced light. To separate this from a Cherenkov event, we look for events which have hits that are spread out in time, since unlike Cherenkov events, flasher events are not necessarily instantaneous. We want to avoid using the full machinery of event reconstruction, so we simply look at the timing of adjacent PMTs. Since we are looking at the hit timing and not merely the position in electronic space, this is one of a few cuts which is sensitive to PMT calibrations and is therefore less robust. This cut works by creating a list of all PMTs which are hit and have good calibration (ignoring the rest), then each hit is compared with every other hit to check for their distance in the detector d and the time difference between their hits Δt . A valid pair is one where,

$$\Delta t < 25ns$$

$$d < 3m.$$

One caveat to this is we want to ignore the PMT hits which are due to pickup in channels near the blind flasher. To do this we ignore any pairs of hits which are due to a cluster (defined as 3

or more channels in a row in electronic space). We then look at the median of the distribution of Δt as long as the number of valid pairs > 15 , and compare this to a median threshold. If $\text{median}[\Delta t] > 6.8 \text{ ns}$, then the event is flagged by the FTS cut.

A.4 Flasher Geometry Cut

With the intent at looking for typical flasher events, the flasher geometry cut looks for a burst of hits nearby (in either electronic or detector space) and compares that to all other hits. Hits are once again compared pair-wise, but in this case we allow for uncalibrated hit information. For each hit, we look at the ratio of all hits within 1 meter to total hits, and if this ratio is greater than 0.5 and the number of hits within the radius is at least 4, then we consider this cluster as a possible flasher origin. For each possible flasher origin, if the average hit time inside the cluster is more than 500 ADC counts ($\sim 48.8 \text{ ns}$) away from the average hit time outside the cluster or if one of the hits in the cluster is considered a bad hitBad hit could be unphysical charge in any of the charge integration channels, and the distance between the cluster and average outer hits is > 12 meters, then the event is flagged. If an event passes this criteria, then another check is made in a similar fashion but clusters are looked for in electronic space. In this case the criteria is that a cluster (at least 4 hits) must be within ± 8 channels of one hit. If a cluster is found in this way, then the same checks for position and time relative to the cluster are made as before.

A.5 In Time Channel Time Spread

The In Time Channel (ITC) Time Spread cut ignores the geometry of events and instead looks only at the timing profile of the hits (which have good calibration). This cut simply slides a 93 ns time window (in 1ns increments) and calculates the percentage of hits within that window. If the window with the highest percentage of hits contains less than 60% of hits, the event is cut from analysis. This primarily targets flat-tac events which have a more uniform distribution of hits across time, whereas Cherenkov events will have most hits clustered in time. Since the flag is based on the time distribution of Cherenkov events, it is an effective flag for water data but will not necessarily be helpful with scintillator.

A.6 Junk

The junk cut is a very simple cut to remove events with anomalous channels being placed within the same event. The cut simply checks that every hit in an event is unique (no PMT contributes more than once), which can happen in the case of an orphaned hit being placed in the wrong event.

A.7 Neck

The neck cut specifically targets neck events using data from the neck PMTs. During the SNO+ water phase we started with 3 PMTs in the neck, and then eventually reduced to only 2 PMTs. In the simplest case, if 2 neck tubes are hit within an event, then that event immediately fails this cut. If only 1 tube is hit, then additional checks are made to look for

the flashlight nature of the acrylic neck, which produces a majority of hits in the bottom half of the detector. If the average time of hits below the equator is more than 85 ns after the neck hit, then the event is flagged.

A.8 OWL

Another simple cut is placed on events in which the number of outward looking PMT hits is greater or equal to 3.

A.9 Charge Cluster

The charge cluster cut is purely in electronics space and looks for high charge hits that cause pickup in surrounding channels (such as wet-end breakdowns, flashers, and sharkfins). A 5 channel wide sliding window (in PMT id) goes across all channels and if 3 of the channels register a hit, then we check if any of the hits are considered bad, and if so the event is flagged.

A.10 QvNhit

The QvNhit cut compares the charge of hits in an event with the total number of hits with the intent in removing events with very little charge in most of the channels. This cut is aimed at low level pickup events which cause many hits just above threshold. Since this cut uses charge information it is important that hits have good calibrations for the cut to be effective. In QvNhit we sum the charge from the lower 90% of hits and then convert this charge into a total number of photoelectrons (based on the calibration for each channel). If the average

number of photoelectrons per channel in this population is less than 0.25, then the event fails.

A.11 QvT

The charge vs time cut is yet another cut aimed at targeting flashers using calibrated PMT data if available. This cut is more relaxed than the flasher geometry cut due to only using charge and timing information. With QvT we look for the hit in an event with the highest charge in two of the three charge channels (for water phase we used QLX and QHL). To qualify as a high charge, we require,

$$QHL_{\max} - hQHL_i \geq 610$$

$$QLX_{\max} - hQLX_i \geq 110.$$

Then for each of these two charge channels we check to see if the hit came earlier than the median by an amount that would indicate a flasher,

$$60 \text{ ns} \leq \text{Median}[\text{time}] - \text{timemax} \leq 250 \text{ ns},$$

and if this is also true, then the event is flagged.

A.12 Ring of Fire

As the name implies, the Ring of Fire cut aims to remove ring of fire events from data. The cut is made using only information in electronics space, so no PMT calibration is needed. To apply the cut, 70% of the hits in the event must be in a single crate, and if so then 80% of those hits must be in the outer ring of the crate. For the water phase we defined the outer ring as cards 0-3 and 12-15 as well as channels 0-7 and 24-31.

In the SNO+ Data Cleaning review in May 2018, it was noted that this cut is redundant when the crateisotropy cut is also implemented. Future SNO+ phases should remove this cut from the analysis mask, but the cut has been left in for the water phase so analyses will not have to be repeated to ensure the ringoffire has no impact.

A.13 CAEN

The CAEN digitizer takes the output of the ESUM trigger signal for the entire event and stores the signal in the data file. This allows us to make cuts on the shape and size of the ESUM trigger based upon what we expect to see for normal events. This has the potential to provide us with high levels of discrimination for various event types (for example, as is shown in the flasher example ??, the CAEN trace has two humps corresponding to the initial flash and the light detected on the other side of the detector). As of the water phase, the only two parameters we are using is the peak height and the integral. Shown in Figure 41 are the bounds in which a good event must satisfy, else we cut that event out of the data. These bounds are chosen to provide a 99% coverage across the 16N calibration run, with corrections made at low hit where pedestal noise becomes comparable to the peak height,

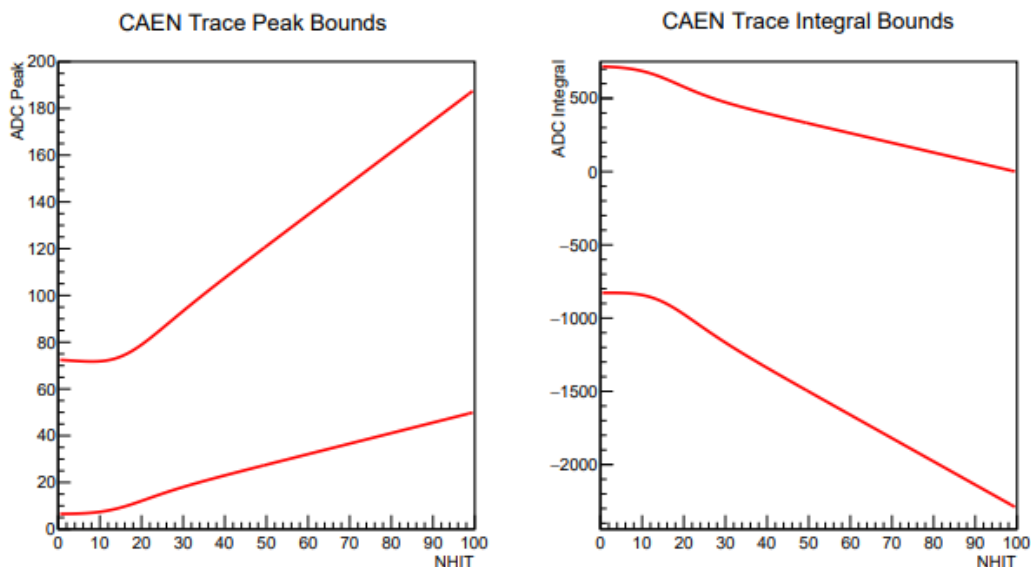
$$a(1 - S(x)) + S(x)(b + cx)$$

where,

$$S(x) = \frac{1}{e^{-(x-TNHIT)/Tw} + 1}$$

	T_{NHIT}	T_w	a	b	c
Peak (lower bound)	15	5	6.6	5.1	0.45
Peak (upper bound)	15	5	73.4	54.14	1.34
Integral (lower bound)	15	5	-834.0	-707.0	-15.9
Integral (upper bound)	15	5	720.6	658.8	-6.61

CAEN Cut constants for SNO+ water phase.



CAEN cut digitized peak (left) and integral (right) bounds as a function of NHIT.

is the Sigmoid function. The Sigmoid is used to parametrize the transition of small amplitude events (low NHIT) into the noise and is tuned using random trigger events (PulseGT in SNO+). T_{NHIT} shifts the Sigmoid to the NHIT at which we transition out of the noise and T_w adjusts the width of that transition. The constants define the linear components of the fit where a controls the baseline at low NHIT, b controls the high NHIT offset, and c is the slope at high NHIT. Notice that outside of the noise range ($S(x) \rightarrow 1$), the boundaries are simply defined by the line $(b + cx)$. For SNO+ water phase the values of these constants are given in Table 4.

A.14 Muon Followers

Due to the depth of SNO+, our muon rate is low enough that we can cut out extremely long windows of time without a significant impact on our livetime. At a muon rate of ~ 3 through the detection volumen per hour, we apply a data cleaning cut to every event for 20 seconds after each muon, resulting in the loss of 1 minute of data each hour. This time window suppresses the vast majority of cosmogenic induced radioisotopes and spallation neutrons except for a few rare candidates (^{11}Be , ^{16}N) which already have tight bounds set thanks to Super-Kamiokande. In the case of a possible cosmogenic follower analysis, an additional data cleaning bit is set tagging events up to 60 seconds after a muon, though this will not be used in physics analysis. The muon tag which this is associated is rather simple (to avoid missing muons) and simply flags any event with at least 5 OWL hits and 150 hits on inward facing PMTs.

A.15 Missed Muon Follower

The missed muon follower looks in some window (Δt) after an event with n_{hit} above some threshold value (N_1) for another event with n_{hit} above some secondary threshold value (N_2). If such an event is found both the initial event and all events with in the Δt window are cut. The ostensible purpose of the cut is to flag events which have a higher likelihood of being muon/atmospheric followed by a neutron. However upon application the best cut parameters did not end up corresponding to those you might expect for cosmogenic neutrons. This suggests another source of time correlated background events that is also flagged by this cut.

For the SNO+ water phase the parameters of the cut are:

- $\Delta t = 1\text{ms}$
- $N_1 = 60$
- $N_2 = 20$

A.16 Polling Cut

There are a few monitoring processes done by the DAQ on a regular basis that involve communicating with the FECs. E.g. CMOS rate polling and sending pedestals. The effect this has on the data taken is not well known and it's feared it might introduce extra noise or result in hits failing to be read out properly. Thus it was determined that the most prudent way to handle these monitoring processes is to impose a dead time while they're going on. In anticipation of this the DAQ typically sets the EXT-2 bit on the MTCD high (using TUBii to do so) while these processes are happening. So the Polling Cut simply looks at the MTCD and TUBii trigger word for each event and if the EXT-2 and MTCA MIMIC1 bit are set in the MTCD and TUBii respectively the event is cut.

A.17 Ped Cut

The Ped Cut serves largely the same purpose as the Polling Cut. It simply cuts for some time window around every pedestal event in data. The nominal value for that time window is 1 second. This cut was added in addition to the Polling Cut for a few reasons, the first and foremost was that for a time in the water phase TUBii was not functioning reliably. The

second reason being that TUBii's standard triggering cannot be used while it's setting the MTCDs EXT-2 high. This was acceptable for the water phase because TUBii wasn't being otherwise used. However, it limits our ability to use it in the future so its beneficial to find an alternative way of flagging and cutting events that take place while the FECs are being accessed.

B Burst File Lists

B.1 Names of bursts in regions 1, 2 and 3

burst_0000105328_000_2.zdab
burst_0000105328_000_48.zdab
burst_0000105328_000_57.zdab
burst_0000105329_000_49.zdab
burst_0000105379_002_129.zdab
burst_0000105379_002_147.zdab
burst_0000105379_002_156.zdab
burst_0000105379_002_22.zdab
burst_0000105379_002_31.zdab
burst_0000105379_002_40.zdab
burst_0000105379_002_4.zdab
burst_0000105379_002_59.zdab
burst_0000105379_002_68.zdab
burst_0000105379_002_77.zdab
burst_0000105379_002_86.zdab
burst_0000105383_000_112.zdab
burst_0000105383_000_121.zdab
burst_0000105383_000_130.zdab
burst_0000105383_000_13.zdab
burst_0000105383_000_158.zdab
burst_0000105383_000_176.zdab
burst_0000105383_000_266.zdab
burst_0000105383_000_31.zdab
burst_0000105383_000_356.zdab
burst_0000105383_000_392.zdab
burst_0000105383_000_400.zdab
burst_0000105383_000_40.zdab
burst_0000105383_000_419.zdab
burst_0000105383_000_428.zdab
burst_0000105383_000_437.zdab
burst_0000105383_000_47.zdab
burst_0000105383_000_74.zdab
burst_0000105383_000_83.zdab
burst_0000105429_001_100.zdab
burst_0000105429_001_119.zdab
burst_0000105429_001_128.zdab
burst_0000105429_001_245.zdab
burst_0000105429_001_254.zdab
burst_0000105429_001_263.zdab
burst_0000105429_001_2.zdab
burst_0000105429_001_39.zdab
burst_0000105450_002_156.zdab
burst_0000105450_002_20.zdab
burst_0000105450_002_228.zdab
burst_0000105450_002_237.zdab
burst_0000105450_002_246.zdab
burst_0000105450_002_255.zdab
burst_0000105450_002_29.zdab
burst_0000105450_002_38.zdab
burst_0000105450_002_65.zdab
burst_0000105450_002_74.zdab
burst_0000105450_002_92.zdab
burst_0000105454_000_100.zdab
burst_0000105454_000_119.zdab
burst_0000105454_000_128.zdab
burst_0000105454_000_137.zdab
burst_0000105454_000_155.zdab
burst_0000105454_000_164.zdab
burst_0000105454_000_182.zdab
burst_0000105454_000_191.zdab
burst_0000105454_000_245.zdab
burst_0000105454_000_371.zdab
burst_0000105509_003_1.zdab
burst_0000105520_002_106.zdab
burst_0000105520_002_115.zdab
burst_0000105520_002_133.zdab
burst_0000105520_002_51.zdab
burst_0000105520_002_60.zdab
burst_0000105520_002_6.zdab
burst_0000105725_001_10.zdab
burst_0000105883_000_1178.zdab
burst_0000105883_000_124.zdab
burst_0000105883_000_1303.zdab
burst_0000105883_000_197.zdab
burst_0000105883_000_205.zdab
burst_0000105883_000_250.zdab
burst_0000105883_000_340.zdab
burst_0000105883_000_476.zdab
burst_0000106207_001_5.zdab
burst_0000106207_001_87.zdab
burst_0000106207_002_105.zdab
burst_0000106207_002_123.zdab
burst_0000106207_002_96.zdab
burst_0000106207_003_32.zdab
burst_0000106207_003_41.zdab
burst_0000106275_000_111.zdab
burst_0000106275_000_148.zdab
burst_0000106275_000_175.zdab
burst_0000106275_000_274.zdab
burst_0000106275_000_337.zdab
burst_0000106275_000_508.zdab
burst_0000106569_001_73.zdab
burst_0000106569_001_82.zdab
burst_0000100393_000_386.zdab
burst_0000100393_000_395.zdab
burst_0000100393_000_403.zdab
burst_0000100393_000_412.zdab
burst_0000100393_000_421.zdab
burst_0000100393_000_430.zdab
burst_0000100393_000_458.zdab
burst_0000100393_000_467.zdab
burst_0000100393_000_476.zdab
burst_0000100393_000_485.zdab
burst_0000100393_000_502.zdab
burst_0000100393_000_511.zdab
burst_0000100393_000_520.zdab
burst_0000100393_000_539.zdab
burst_0000100393_000_548.zdab
burst_0000100393_000_557.zdab
burst_0000100393_000_566.zdab
burst_0000100393_000_575.zdab
burst_0000100393_000_584.zdab
burst_0000100393_000_593.zdab
burst_0000100393_000_601.zdab
burst_0000100393_000_610.zdab
burst_0000100393_000_638.zdab
burst_0000100393_000_647.zdab
burst_0000100393_000_656.zdab
burst_0000100393_000_665.zdab
burst_0000100393_000_674.zdab
burst_0000100393_000_683.zdab
burst_0000100393_000_692.zdab
burst_0000100393_000_700.zdab
burst_0000100393_000_719.zdab
burst_0000100393_000_737.zdab
burst_0000100393_000_746.zdab
burst_0000100393_000_755.zdab
burst_0000100393_000_764.zdab
burst_0000100393_000_773.zdab
burst_0000100393_000_782.zdab
burst_0000100393_000_791.zdab
burst_0000100393_000_7.zdab
burst_0000100393_000_809.zdab
burst_0000100393_000_818.zdab
burst_0000100393_000_827.zdab
burst_0000100393_000_836.zdab
burst_0000100393_000_854.zdab
burst_0000100393_000_863.zdab
burst_0000100393_000_881.zdab
burst_0000100393_000_890.zdab
burst_0000100393_000_89.zdab
burst_0000100393_000_908.zdab
burst_0000100393_000_917.zdab
burst_0000100393_000_935.zdab
burst_0000100393_000_944.zdab
burst_0000100393_000_953.zdab
burst_0000100393_000_962.zdab
burst_0000100393_000_971.zdab
burst_0000100393_000_980.zdab
burst_0000100393_000_98.zdab
burst_0000100393_001_106.zdab
burst_0000100393_001_115.zdab
burst_0000100393_001_30.zdab
burst_0000100393_001_3.zdab
burst_0000100393_001_49.zdab
burst_0000100393_001_67.zdab
burst_0000100393_001_76.zdab
burst_0000100393_001_85.zdab
burst_0000100393_001_94.zdab
burst_0000100549_000_1099.zdab
burst_0000100549_000_1206.zdab
burst_0000100549_000_1297.zdab
burst_0000100549_000_1639.zdab
burst_0000100549_000_303.zdab
burst_0000100549_000_691.zdab
burst_0000100549_000_826.zdab
burst_0000100549_000_835.zdab
burst_0000100549_000_844.zdab
burst_0000100549_000_943.zdab
burst_0000100549_001_1041.zdab
burst_0000100549_001_1087.zdab
burst_0000100549_001_1122.zdab
burst_0000100549_001_148.zdab
burst_0000100549_001_247.zdab
burst_0000100549_001_265.zdab
burst_0000100549_001_841.zdab
burst_0000100550_001_21.zdab
burst_0000100550_001_6.zdab
burst_0000100551_000_114.zdab
burst_0000100551_000_204.zdab
burst_0000100551_000_231.zdab
burst_0000100551_000_484.zdab
burst_0000100551_001_1000.zdab
burst_0000100551_001_1046.zdab
burst_0000100551_001_1055.zdab
burst_0000100551_001_1109.zdab

burst_0000108372_000_528.zdab
burst_0000108372_000_564.zdab
burst_0000108372_000_80.zdab
burst_0000108533_000_187.zdab
burst_0000108533_000_268.zdab
burst_0000108533_000_295.zdab
burst_0000109080_000_0.zdab
burst_0000109091_001_195.zdab
burst_0000109091_002_614.zdab
burst_0000109091_002_623.zdab
burst_0000109091_002_632.zdab
burst_0000109091_002_641.zdab
burst_0000109091_002_650.zdab
burst_0000109091_002_768.zdab
burst_0000109091_002_867.zdab
burst_0000109091_002_876.zdab
burst_0000109091_002_885.zdab
burst_0000109091_002_894.zdab
burst_0000109091_002_902.zdab
burst_0000109091_002_911.zdab
burst_0000109091_002_92.zdab
burst_0000109091_002_984.zdab
burst_0000109091_002_993.zdab
burst_0000109126_000_5.zdab
burst_0000109414_000_35.zdab
burst_0000109546_000_1.zdab

burst_0000101307_001_14.zdab
burst_0000101307_001_23.zdab
burst_0000101307_001_41.zdab
burst_0000101307_001_50.zdab
burst_0000101307_001_5.zdab
burst_0000101307_001_69.zdab
burst_0000101307_001_78.zdab
burst_0000101645_003_15.zdab
burst_0000101645_004_11.zdab
burst_0000102008_000_0.zdab
burst_0000102032_000_109.zdab
burst_0000102032_000_118.zdab
burst_0000102032_000_127.zdab
burst_0000102032_001_30.zdab
burst_0000102035_003_229.zdab
burst_0000102035_003_247.zdab
burst_0000102035_003_274.zdab
burst_0000102035_003_292.zdab
burst_0000102035_003_300.zdab
burst_0000102035_003_30.zdab
burst_0000102035_003_57.zdab
burst_0000102035_003_66.zdab
burst_0000102035_003_75.zdab
burst_0000102035_003_84.zdab
burst_0000102035_003_93.zdab
burst_0000102235_002_30.zdab
burst_0000102235_002_328.zdab
burst_0000102235_002_337.zdab
burst_0000102235_002_57.zdab
burst_0000102235_002_66.zdab
burst_0000102235_002_93.zdab
burst_0000102236_001_153.zdab
burst_0000102236_001_162.zdab
burst_0000102236_001_171.zdab
burst_0000102237_003_31.zdab
burst_0000102237_003_329.zdab
burst_0000103676_000_25.zdab
burst_0000103676_000_52.zdab
burst_0000103686_000_9.zdab
burst_0000103763_000_53.zdab
burst_0000103980_000_125.zdab
burst_0000103980_000_134.zdab
burst_0000103980_001_128.zdab
burst_0000103980_001_209.zdab
burst_0000103980_001_245.zdab
burst_0000103980_001_55.zdab
burst_0000103980_001_64.zdab
burst_0000103980_001_82.zdab
burst_0000103980_001_91.zdab
burst_0000104089_000_0.zdab
burst_0000104935_003_101.zdab
burst_0000104935_003_110.zdab
burst_0000104935_003_147.zdab
burst_0000104935_003_165.zdab
burst_0000104935_003_55.zdab
burst_0000104938_000_27.zdab
burst_0000105061_000_161.zdab
burst_0000105061_000_17.zdab
burst_0000105061_000_242.zdab
burst_0000105061_000_84.zdab
burst_0000105081_001_0.zdab
burst_0000105081_001_63.zdab
burst_0000105081_001_72.zdab
burst_0000105083_001_398.zdab
burst_0000105086_001_341.zdab
burst_0000105086_001_404.zdab
burst_0000105086_001_440.zdab
burst_0000105094_000_146.zdab
burst_0000105094_001_54.zdab
burst_0000105094_001_72.zdab
burst_0000105094_001_90.zdab
burst_0000105095_001_202.zdab
burst_0000105095_001_301.zdab
burst_0000105095_001_31.zdab
burst_0000105097_000_95.zdab
burst_0000105099_003_4.zdab
burst_0000105099_003_68.zdab
burst_0000105100_001_249.zdab
burst_0000105100_001_285.zdab
burst_0000105100_001_294.zdab
burst_0000105100_001_302.zdab
burst_0000105100_001_375.zdab
burst_0000105100_001_72.zdab
burst_0000105101_003_52.zdab
burst_0000105101_003_61.zdab
burst_0000105103_000_188.zdab
burst_0000105103_000_223.zdab
burst_0000105103_000_241.zdab
burst_0000105103_000_250.zdab
burst_0000105103_000_38.zdab
burst_0000105103_000_92.zdab
burst_0000105104_000_146.zdab
burst_0000105104_000_164.zdab

burst_0000105104_000_182.zdab
burst_0000105108_000_114.zdab
burst_0000105108_000_240.zdab
burst_0000105108_000_51.zdab
burst_0000105108_000_60.zdab
burst_0000105108_000_88.zdab
burst_0000105109_002_240.zdab
burst_0000105109_002_268.zdab
burst_0000105109_002_394.zdab
burst_0000105109_002_411.zdab
burst_0000105109_002_67.zdab
burst_0000105115_001_252.zdab
burst_0000105117_002_219.zdab
burst_0000105147_000_6.zdab
burst_0000105147_000_97.zdab
burst_0000105221_001_122.zdab
burst_0000105221_001_131.zdab
burst_0000105221_001_140.zdab
burst_0000105221_001_46.zdab
burst_0000105221_001_91.zdab
burst_0000105223_001_0.zdab
burst_0000105223_001_316.zdab
burst_0000105223_001_343.zdab
burst_0000105223_001_352.zdab
burst_0000105223_001_406.zdab
burst_0000105226_000_92.zdab
burst_0000105270_001_150.zdab
burst_0000105270_001_59.zdab
burst_0000105270_001_68.zdab
burst_0000105271_001_239.zdab
burst_0000105271_001_293.zdab
burst_0000105279_000_178.zdab
burst_0000105316_001_251.zdab
burst_0000105316_001_297.zdab
burst_0000105316_001_305.zdab
burst_0000105316_001_59.zdab
burst_0000105318_000_4.zdab
burst_0000105323_001_140.zdab
burst_0000105323_001_186.zdab
burst_0000105323_001_267.zdab
burst_0000105323_001_276.zdab
burst_0000105323_001_91.zdab
burst_0000105328_000_66.zdab
burst_0000105329_000_30.zdab
burst_0000105329_000_67.zdab
burst_0000105379_002_101.zdab
burst_0000105379_002_138.zdab
burst_0000105379_002_95.zdab
burst_0000105383_000_103.zdab
burst_0000105383_000_149.zdab
burst_0000105383_000_167.zdab
burst_0000105383_000_185.zdab
burst_0000105383_000_194.zdab
burst_0000105383_000_202.zdab
burst_0000105383_000_211.zdab
burst_0000105383_000_220.zdab
burst_0000105383_000_22.zdab
burst_0000105383_000_248.zdab
burst_0000105383_000_257.zdab
burst_0000105383_000_275.zdab
burst_0000105383_000_284.zdab
burst_0000105383_000_293.zdab
burst_0000105383_000_301.zdab
burst_0000105383_000_310.zdab
burst_0000105383_000_329.zdab
burst_0000105383_000_338.zdab
burst_0000105383_000_383.zdab
burst_0000105429_001_272.zdab
burst_0000105429_001_281.zdab
burst_0000105429_001_48.zdab
burst_0000105429_001_57.zdab
burst_0000105429_001_66.zdab
burst_0000105450_002_219.zdab
burst_0000105450_002_56.zdab
burst_0000105454_000_173.zdab
burst_0000105454_000_1.zdab
burst_0000105454_000_335.zdab
burst_0000105454_000_380.zdab
burst_0000105520_002_33.zdab
burst_0000105520_002_42.zdab
burst_0000105812_002_5.zdab
burst_0000105844_003_4.zdab
burst_0000105846_003_16.zdab
burst_0000105848_000_2.zdab
burst_0000105883_000_1006.zdab
burst_0000105883_000_1015.zdab
burst_0000105883_000_1024.zdab
burst_0000105883_000_1033.zdab
burst_0000105883_000_1042.zdab
burst_0000105883_000_1051.zdab
burst_0000105883_000_1060.zdab
burst_0000105883_000_106.zdab
burst_0000105883_000_1079.zdab

burst_0000105883_000_1088.zdab
burst_0000105883_000_1097.zdab
burst_0000105883_000_1105.zdab
burst_0000105883_000_1114.zdab
burst_0000105883_000_1141.zdab
burst_0000105883_000_1150.zdab
burst_0000105883_000_115.zdab
burst_0000105883_000_1169.zdab
burst_0000105883_000_1187.zdab
burst_0000105883_000_1196.zdab
burst_0000105883_000_1204.zdab
burst_0000105883_000_1213.zdab
burst_0000105883_000_1222.zdab
burst_0000105883_000_1231.zdab
burst_0000105883_000_1240.zdab
burst_0000105883_000_1259.zdab
burst_0000105883_000_1268.zdab
burst_0000105883_000_1277.zdab
burst_0000105883_000_1286.zdab
burst_0000105883_000_1295.zdab
burst_0000105883_000_1312.zdab
burst_0000105883_000_142.zdab
burst_0000105883_000_151.zdab
burst_0000105883_000_160.zdab
burst_0000105883_000_16.zdab
burst_0000105883_000_179.zdab
burst_0000105883_000_188.zdab
burst_0000105883_000_214.zdab
burst_0000105883_000_232.zdab
burst_0000105883_000_241.zdab
burst_0000105883_000_25.zdab
burst_0000105883_000_269.zdab
burst_0000105883_000_278.zdab
burst_0000105883_000_287.zdab
burst_0000105883_000_296.zdab
burst_0000105883_000_304.zdab
burst_0000105883_000_313.zdab
burst_0000105883_000_322.zdab
burst_0000105883_000_331.zdab
burst_0000105883_000_359.zdab
burst_0000105883_000_368.zdab
burst_0000105883_000_377.zdab
burst_0000105883_000_386.zdab
burst_0000105883_000_403.zdab
burst_0000105883_000_412.zdab
burst_0000105883_000_421.zdab
burst_0000105883_000_430.zdab
burst_0000105883_000_43.zdab
burst_0000105883_000_449.zdab
burst_0000105883_000_458.zdab
burst_0000105883_000_467.zdab
burst_0000105883_000_485.zdab
burst_0000105883_000_494.zdab
burst_0000105883_000_502.zdab
burst_0000105883_000_511.zdab
burst_0000105883_000_520.zdab
burst_0000105883_000_548.zdab
burst_0000105883_000_557.zdab
burst_0000105883_000_566.zdab
burst_0000105883_000_584.zdab
burst_0000105883_000_593.zdab
burst_0000105883_000_601.zdab
burst_0000105883_000_610.zdab
burst_0000105883_000_61.zdab
burst_0000105883_000_629.zdab
burst_0000105883_000_647.zdab
burst_0000105883_000_665.zdab
burst_0000105883_000_674.zdab
burst_0000105883_000_683.zdab
burst_0000105883_000_692.zdab
burst_0000105883_000_700.zdab
burst_0000105883_000_70.zdab
burst_0000105883_000_719.zdab
burst_0000105883_000_728.zdab
burst_0000105883_000_755.zdab
burst_0000105883_000_764.zdab
burst_0000105883_000_773.zdab
burst_0000105883_000_782.zdab
burst_0000105883_000_791.zdab
burst_0000105883_000_7.zdab
burst_0000105883_000_809.zdab
burst_0000105883_000_818.zdab
burst_0000105883_000_827.zdab
burst_0000105883_000_836.zdab
burst_0000105883_000_854.zdab
burst_0000105883_000_863.zdab
burst_0000105883_000_872.zdab
burst_0000105883_000_881.zdab
burst_0000105883_000_890.zdab
burst_0000105883_000_89.zdab
burst_0000105883_000_908.zdab
burst_0000105883_000_917.zdab
burst_0000105883_000_926.zdab

burst_0000105883_000_935.zdab
burst_0000105883_000_944.zdab
burst_0000106123_001_23.zdab
burst_0000106123_001_5.zdab
burst_0000106207_001_102.zdab
burst_0000106207_001_111.zdab
burst_0000106207_001_14.zdab
burst_0000106207_001_23.zdab
burst_0000106207_001_32.zdab
burst_0000106207_001_96.zdab
burst_0000106207_002_114.zdab
burst_0000106207_002_132.zdab
burst_0000106207_002_141.zdab
burst_0000106207_002_150.zdab
burst_0000106207_002_15.zdab
burst_0000106207_002_169.zdab
burst_0000106207_002_23.zdab
burst_0000106207_002_32.zdab
burst_0000106207_002_41.zdab
burst_0000106207_002_69.zdab
burst_0000106207_002_78.zdab
burst_0000106207_003_50.zdab
burst_0000106207_003_5.zdab
burst_0000106275_000_102.zdab
burst_0000106275_000_120.zdab
burst_0000106275_000_12.zdab
burst_0000106275_000_166.zdab
burst_0000106275_000_193.zdab
burst_0000106275_000_201.zdab
burst_0000106275_000_21.zdab
burst_0000106275_000_229.zdab
burst_0000106275_000_238.zdab
burst_0000106275_000_247.zdab
burst_0000106275_000_256.zdab
burst_0000106275_000_265.zdab
burst_0000106275_000_283.zdab
burst_0000106275_000_292.zdab
burst_0000106275_000_300.zdab
burst_0000106275_000_30.zdab
burst_0000106275_000_319.zdab
burst_0000106275_000_328.zdab
burst_0000106275_000_346.zdab
burst_0000106275_000_355.zdab
burst_0000106275_000_364.zdab
burst_0000106275_000_3.zdab
burst_0000106275_000_445.zdab
burst_0000106275_000_454.zdab
burst_0000106275_000_463.zdab
burst_0000106275_000_472.zdab
burst_0000106275_000_481.zdab
burst_0000106275_000_544.zdab
burst_0000106570_000_44.zdab
burst_0000106570_000_53.zdab
burst_0000106782_000_7.zdab
burst_0000106800_000_16.zdab
burst_0000106800_000_25.zdab
burst_0000106800_000_70.zdab
burst_0000106981_001_104.zdab
burst_0000106981_001_122.zdab
burst_0000106981_001_140.zdab
burst_0000106981_001_14.zdab
burst_0000106981_001_159.zdab
burst_0000106981_001_168.zdab
burst_0000106981_001_177.zdab
burst_0000106981_001_186.zdab
burst_0000106981_001_195.zdab
burst_0000106981_001_203.zdab
burst_0000106981_001_212.zdab
burst_0000106981_001_221.zdab
burst_0000106981_001_23.zdab
burst_0000106981_001_249.zdab
burst_0000106981_001_258.zdab
burst_0000106981_001_276.zdab
burst_0000106981_001_285.zdab
burst_0000106981_001_294.zdab
burst_0000107463_000_235.zdab
burst_0000107463_000_52.zdab
burst_0000107466_000_223.zdab
burst_0000107466_001_12.zdab
burst_0000107471_000_41.zdab
burst_0000107471_000_50.zdab
burst_0000107471_000_5.zdab
burst_0000107471_000_78.zdab
burst_0000107471_000_96.zdab
burst_0000107530_003_104.zdab
burst_0000107530_003_122.zdab
burst_0000107530_003_131.zdab
burst_0000107530_003_140.zdab
burst_0000107530_003_14.zdab
burst_0000107530_003_159.zdab
burst_0000107530_003_177.zdab
burst_0000107530_003_186.zdab
burst_0000107530_003_203.zdab

burst_0000107530_003_221.zdab
burst_0000107530_003_23.zdab
burst_0000107530_003_249.zdab
burst_0000107530_003_258.zdab
burst_0000108372_000_32.zdab
burst_0000108372_000_41.zdab
burst_0000108372_000_537.zdab
burst_0000108372_000_555.zdab
burst_0000108475_002_100.zdab
burst_0000108476_001_1131.zdab
burst_0000108476_001_440.zdab
burst_0000108477_000_269.zdab
burst_0000108477_000_368.zdab
burst_0000108477_000_61.zdab
burst_0000108510_000_35.zdab
burst_0000108510_000_62.zdab
burst_0000108510_000_71.zdab
burst_0000108510_000_80.zdab
burst_0000108531_000_35.zdab
burst_0000108531_000_53.zdab
burst_0000108531_000_6.zdab
burst_0000108533_000_123.zdab
burst_0000108533_000_132.zdab
burst_0000108533_000_169.zdab
burst_0000108533_000_178.zdab
burst_0000108533_000_231.zdab
burst_0000108533_000_277.zdab
burst_0000108533_000_286.zdab
burst_0000108533_000_321.zdab
burst_0000108533_000_33.zdab
burst_0000108533_000_349.zdab
burst_0000108533_000_420.zdab
burst_0000108533_000_42.zdab
burst_0000108533_000_439.zdab
burst_0000108533_000_448.zdab
burst_0000108545_002_904.zdab
burst_0000108545_002_913.zdab
burst_0000108545_002_922.zdab
burst_0000108545_002_931.zdab
burst_0000108545_002_940.zdab
burst_0000108545_002_94.zdab
burst_0000108545_002_959.zdab
burst_0000108545_002_968.zdab
burst_0000108545_002_977.zdab
burst_0000108545_002_986.zdab
burst_0000108545_002_995.zdab
burst_0000108545_003_13.zdab
burst_0000108545_003_22.zdab
burst_0000108545_003_31.zdab
burst_0000108545_003_40.zdab
burst_0000108545_003_59.zdab
burst_0000108553_000_1.zdab
burst_0000108716_000_5.zdab
burst_0000108948_001_13.zdab
burst_0000108948_001_31.zdab
burst_0000108948_001_40.zdab
burst_0000108948_001_4.zdab
burst_0000108948_002_2.zdab
burst_0000109091_001_104.zdab
burst_0000109091_001_140.zdab
burst_0000109091_001_14.zdab
burst_0000109091_001_168.zdab
burst_0000109091_001_177.zdab
burst_0000109091_001_186.zdab
burst_0000109091_001_203.zdab
burst_0000109091_001_212.zdab
burst_0000109091_001_221.zdab
burst_0000109091_001_230.zdab
burst_0000109091_001_249.zdab
burst_0000109091_001_320.zdab
burst_0000109091_001_32.zdab
burst_0000109091_001_339.zdab
burst_0000109091_001_348.zdab
burst_0000109091_001_366.zdab
burst_0000109091_001_375.zdab
burst_0000109091_001_384.zdab
burst_0000109091_001_393.zdab
burst_0000109091_001_401.zdab
burst_0000109091_001_410.zdab
burst_0000109091_001_47.zdab
burst_0000109091_001_65.zdab
burst_0000109091_001_74.zdab
burst_0000109091_001_83.zdab
burst_0000109091_001_92.zdab
burst_0000109091_002_1000.zdab
burst_0000109091_002_100.zdab
burst_0000109091_002_1037.zdab
burst_0000109091_002_1046.zdab
burst_0000109091_002_1055.zdab
burst_0000109091_002_119.zdab
burst_0000109091_002_128.zdab
burst_0000109091_002_137.zdab
burst_0000109091_002_146.zdab

burst_0000109091_002_155.zdab
burst_0000109091_002_164.zdab
burst_0000109091_002_173.zdab
burst_0000109091_002_182.zdab
burst_0000109091_002_191.zdab
burst_0000109091_002_1.zdab
burst_0000109091_002_209.zdab
burst_0000109091_002_218.zdab
burst_0000109091_002_227.zdab
burst_0000109091_002_236.zdab
burst_0000109091_002_245.zdab
burst_0000109091_002_254.zdab
burst_0000109091_002_263.zdab
burst_0000109091_002_272.zdab
burst_0000109091_002_281.zdab
burst_0000109091_002_290.zdab
burst_0000109091_002_29.zdab
burst_0000109091_002_308.zdab
burst_0000109091_002_317.zdab
burst_0000109091_002_326.zdab
burst_0000109091_002_335.zdab
burst_0000109091_002_344.zdab
burst_0000109091_002_353.zdab
burst_0000109091_002_362.zdab
burst_0000109091_002_371.zdab
burst_0000109091_002_380.zdab
burst_0000109091_002_38.zdab
burst_0000109091_002_399.zdab
burst_0000109091_002_407.zdab
burst_0000109091_002_425.zdab
burst_0000109091_002_434.zdab
burst_0000109091_002_443.zdab
burst_0000109091_002_452.zdab
burst_0000109091_002_461.zdab
burst_0000109091_002_470.zdab
burst_0000109091_002_47.zdab
burst_0000109091_002_489.zdab
burst_0000109091_002_515.zdab
burst_0000109091_002_533.zdab
burst_0000109091_002_560.zdab
burst_0000109091_002_669.zdab
burst_0000109091_002_678.zdab
burst_0000109091_002_687.zdab
burst_0000109091_002_696.zdab
burst_0000109091_002_704.zdab
burst_0000109091_002_722.zdab
burst_0000109091_002_731.zdab
burst_0000109091_002_740.zdab
burst_0000109091_002_849.zdab
burst_0000109091_002_966.zdab
burst_0000109091_002_975.zdab
burst_0000109121_000_1.zdab
burst_0000109198_001_7.zdab
burst_0000109375_002_15.zdab
burst_0000109375_002_33.zdab
burst_0000109375_002_5.zdab
burst_0000109378_000_6.zdab
burst_0000109546_000_10.zdab
burst_0000109546_000_29.zdab
burst_0000109546_000_65.zdab
burst_0000109546_000_83.zdab
burst_0000109547_000_3.zdab

B.2 List of burst causes for regions 1, 2 and 3

Run Number	Region 1	Region 2	Region 3
100102		Crate 12 breakdown	
100139	Crate 12 breakdown	Crate 12 breakdown	
100321	Detector ramping		
100352	Light leak	Light leak	
100354	Light leak	Light leak	Light leak
100355	Light leak	Light leak	Light leak
100382	Light leak	Light leak	
100392	Light leak	Light leak	Light leak
100393		Light leak	Light leak
100397			
100549		Ramping detector	Ramping detector
100550		Gate valves opened	
100551	Manipulight	Manipulight	Manipulight
100956		Electrical Maintenance	
101158	Use of radios	Use of radios	Use of radios
101171			

101172			
101256			
101272			
101307	Crate 12 breakdown	Crate 12 breakdown	
101315			
101475			
101525			
101528			
101645	Good run	Good run	
102008		Electrical work	
102032	Baseline shift	Baseline shift	Baseline shift
102035	Baseline shift	Baseline shift	
102036		Baseline shift	
102037	Baseline shift	Baseline shift	
102189			
102216	Timing Errors		
102235	Baseline shift	Baseline shift	
102236	Baseline shift	Baseline shift	
102237	Baseline shift	Baseline shift	
103271			
103315	Timing errors		

103670	Electrical diagnostics		
103673			
103676	Diagnostic running	Diagnostic running	
103686	Diagnostic running	Diagnostic running	
103688			Diagnostic running
103691	Diagnostic running		
103698			
103704			
103706	Detector ramping		
103740	Diagnostic running		
103762	Good run		
103763	Good run	Good run	
103764	Good run		
103816			
103836	Diagnostic running		
103980	Crate 11 Breakdown	Crate 11 breakdown	Crate 11 breakdown
103981	Detector ramping		
104089	Timing errors	Timing errors	
104099	Crate 2 builder		
104399			
104632	Timing error		

104935	Timing error	Timing error	
104938	Timing error	Timing error	
104951	Shining light into AV		
104980	Diagnostic running		
103315	Timing errors		
105016	Timing errors		
105029			
105061	Timing errors	Timing errors	
105076			
105081	Baseline shift	Baseline shift	
105083	Baseline shift	Baseline shift	Baseline shift
105086	Significant breakdown	Significant breakdown	
105094	Timing error	Timing error	
105095	Timing error	Timing error	
105096			
105097	Timing error	Timing error	
105099	Baseline shift	Baseline shift	
105100	Baseline shift	Baseline shift	
105101	Baseline shift	Baseline shift	
105103	Baseline shift	Baseline shift	
105104	Baseline shift	Baseline shift	

105105	Baseline shift	Baseline shift	
105108	Baseline shift	Baseline shift	
105109	Baseline shift	Baseline shift	
105111	Baseline shift		Baseline shift
105115	Baseline shift	Baseline shift	
105117	Baseline shift	Baseline shift	
105119	Baseline shift		Baseline shift
105139	Baseline shift		
105147	Timing errors	Timing errors	
105218			
105221	Baseline shift	Baseline shift	Baseline shift
105223	Baseline shift	Baseline shift	Baseline shift
105225			
105226	Baseline shift	Baseline shift	
105233			
105255			
105256			
105257	Baseline shift	Baseline shift	Baseline shift
105262	Baseline shift		
105268	Baseline shift		
105270	Timing Error	Timing error	

105271	Timing Error	Timing error	
105273			
105277			
105279	Baseline shift	Baseline shift	
105288	Baseline shift		
105290	Baseline shift		
105316	Baseline shift	Baseline shift	Baseline shift
105318	Baseline shift	Baseline shift	
105323	Baseline shift	Baseline shift	
105327			
105328	Baseline shift	Baseline shift	
105329	Baseline shift	Baseline shift	Baseline shift
105379	Timing error	Timing error	
105383	HV Breakdown	HV Breakdown	HV Breakdown
105429	Baseline shift	Baseline shift	
105450	Baseline shift	Baseline shift	
105454	Baseline shift	Baseline shift	Baseline shift
105460			
105509			
105520	Baseline shift	Baseline shift	Baseline shift
105725	Timing Error		

105771			
105812		Timing error	
105844		Timing error	
105846		Timing error	
105848		Timing error	
105883	Calibration induced	Calibration induced	Calibration induced
106123	High rate	High rate	
106207	Timing error	Timing error	
106275	HV Breakdown	HV Breakdown	HV Breakdown
106569	Timing error	Timing error	Timing error
106570	Timing error	Timing error	
106672	Timing rack work		
106781	Timing error		
106782		LV/ramping	
106800	Baseline shift	Baseline shift	Baseline shift
106981	HV Breakdown	HV Breakdown	
107438	Baseline shift		
107463	Baseline shift	Baseline shift	Baseline shift
107466	Baseline shift	Baseline shift	
107471	Baseline shift	Baseline shift	
107530	Laser Maintenance	Laser maintenance	Laser maintenance

107567			
107618			
107952			
108241			
108242	Following Breakdown		
108303	Cavity circulation	Cavity circulation	
108305			
108372	Timing error	Timing error	
108419			
108421			
108422			
108426			
108468			
108475		Timing/trigger error	Timing/trigger error
108476		Timing/trigger error	Timing/trigger error
108477		Timing/trigger error	Timing/trigger error
108510		Timing/trigger error	Timing/trigger error
108531		Timing error	
108533	Timing error	Timing error	Timing error
108542			
108544			

108545		Diagnostic running	
108553		Wet breakdown	
108716		Rate spikes	
108948		HV Breakdown	
109080	Laserball transition		
109091	Calibration induced	Calibration induced	Calibration induced
109121		Swapping sources	
109126	Calibration induced		
109198		Normal operation	
109375		Timing rack reset	
109378		Low voltage	
109380			Detector Ramping
109410			
109411			
109414	AV water removal		AV water removal
108418			Electronics work
109420			
109499			
109505			
109506			
109545			

109546	Electrical Diagnostics	Electrical diagnostics
109547		Electrical diagnostics
109559		
109560		
109566		
109576		
109592		
109593		
109594		
109596		
109597		
109598		

C Additional PCA Plots

C.1 Entire Data set PCA plots

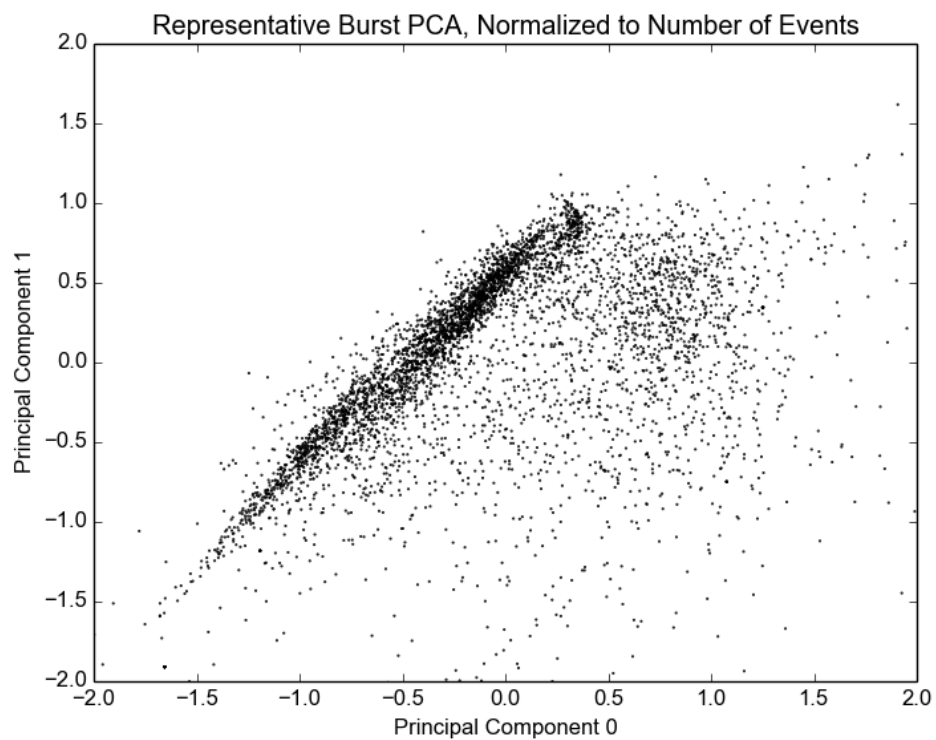


Figure 55: Principal component 1 vs 0, representative burst sample spanning 5180 bursts. 4 visible concentrations of data points.

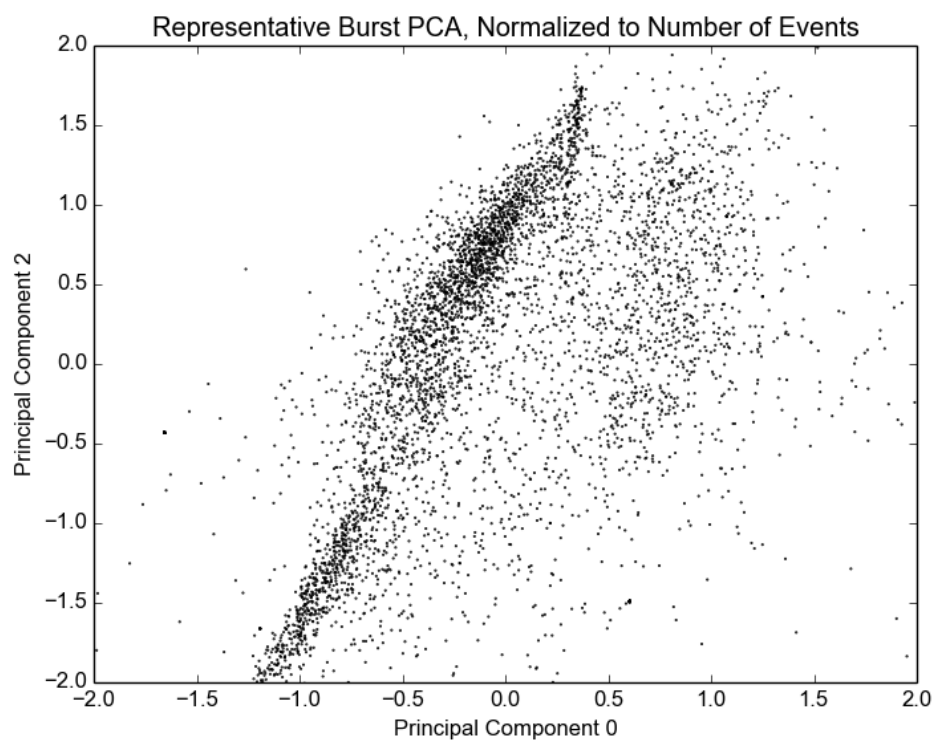


Figure 56: Principal component 2 vs 0, representative burst sample spanning 5180 bursts. 4 visible concentrations of data points.

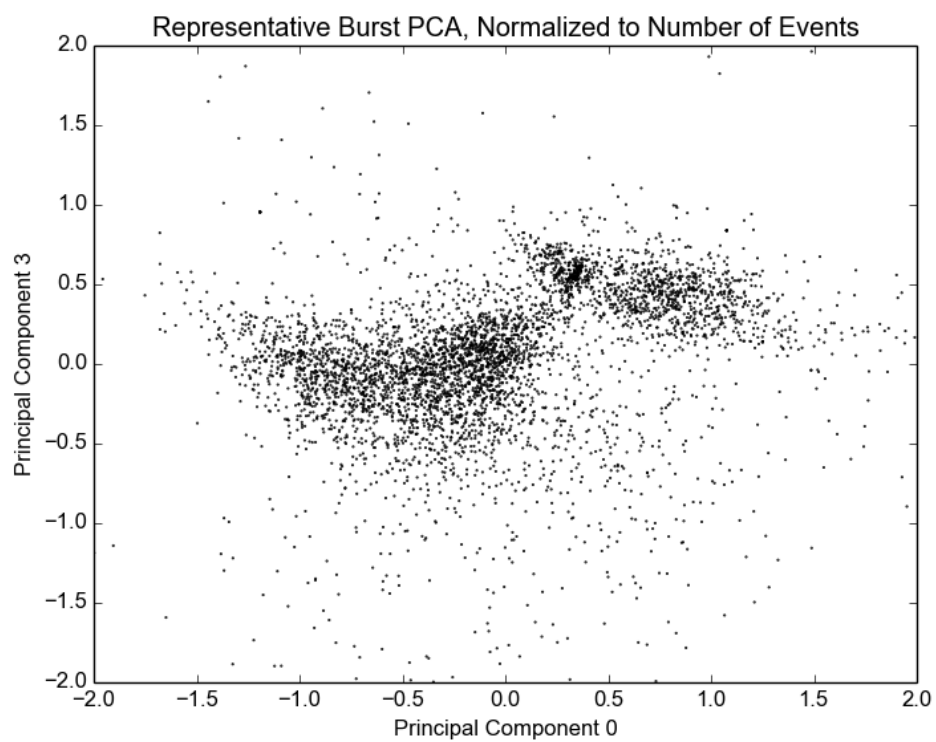


Figure 57: Principal component 3 vs 0, representative burst sample spanning 5180 bursts. 3 visible concentrations of data points.

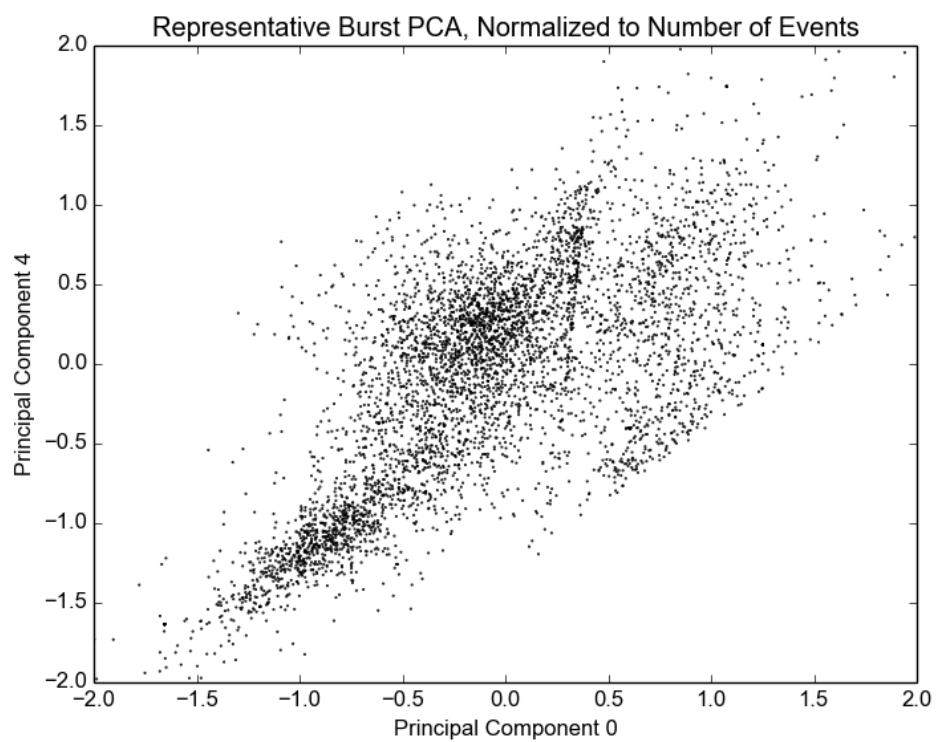


Figure 58: Principal component 4 vs 0, representative burst sample spanning 5180 bursts. 4 visible concentrations of data points.

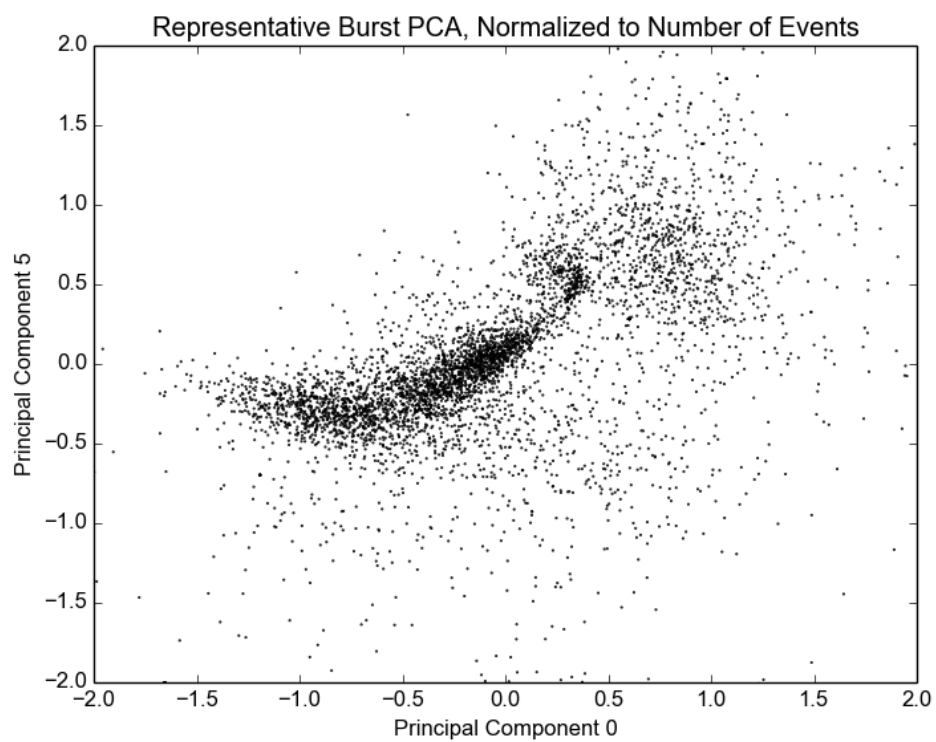


Figure 59: Principal component 5 vs 0, representative burst sample spanning 5180 bursts. 3 visible concentrations of data points.

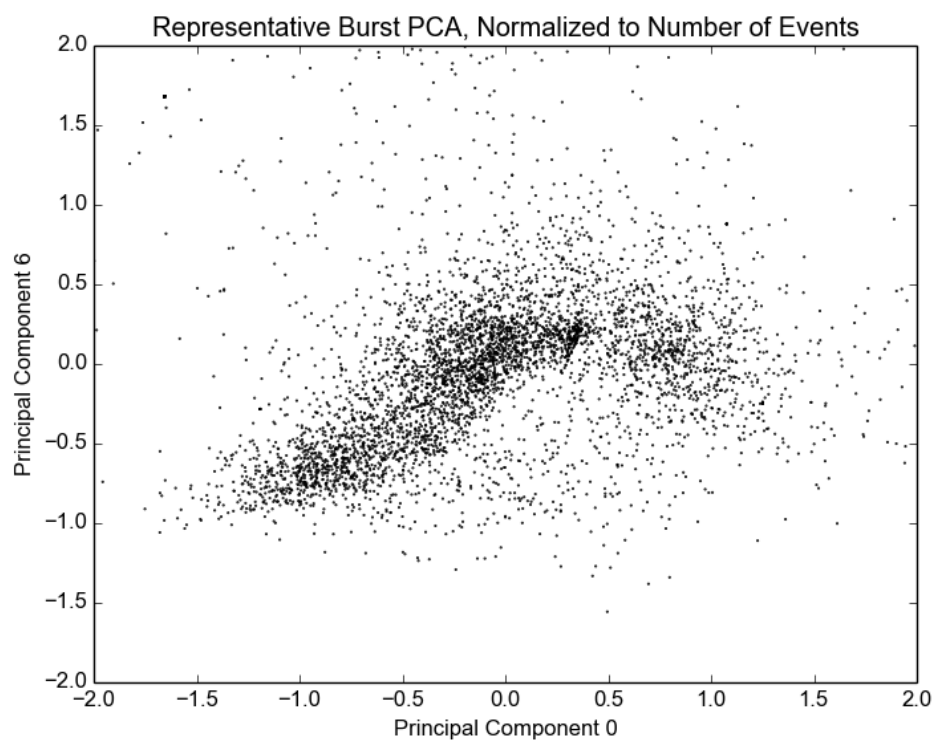


Figure 60: Principal component 6 vs 0, representative burst sample spanning 5180 bursts. 4 visible concentrations of data points.

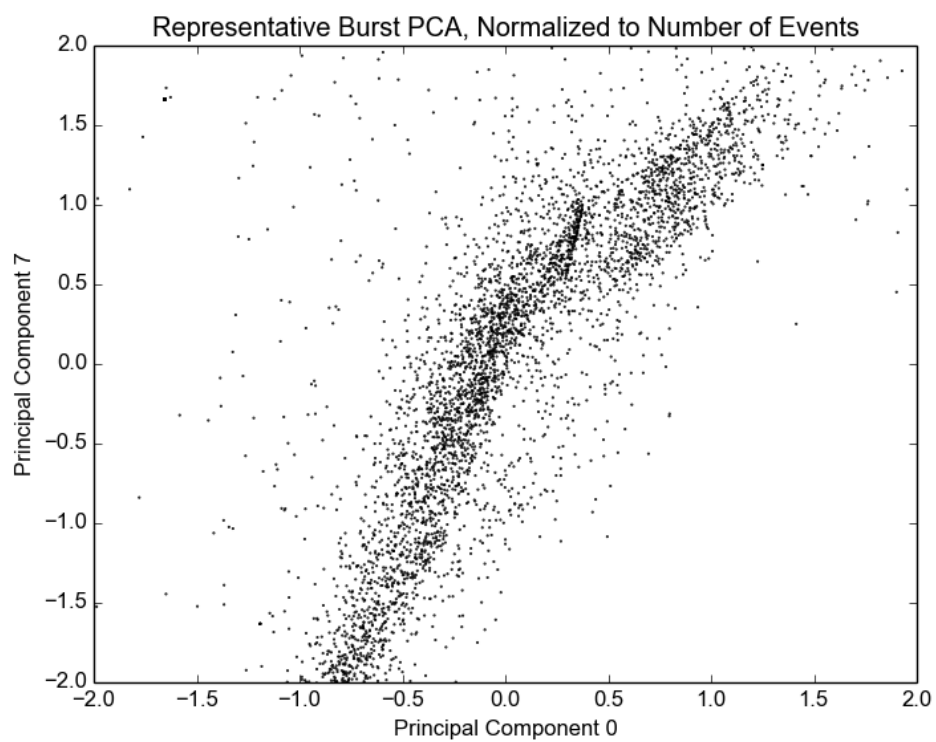


Figure 61: Principal component 7 vs 0, representative burst sample spanning 5180 bursts. Shows a regional distribution of data points.

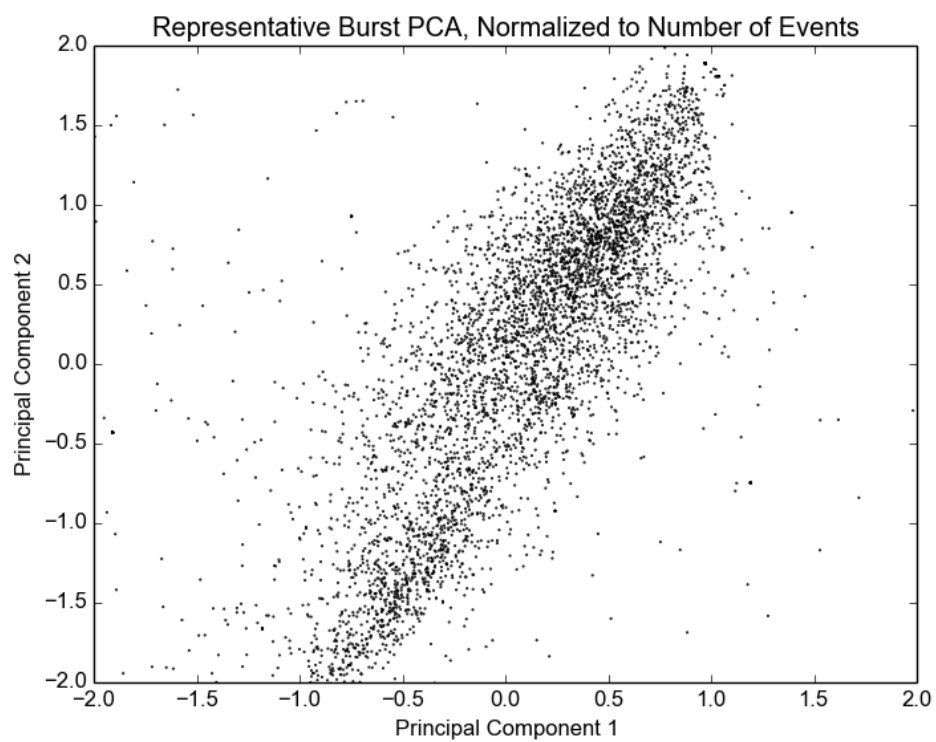


Figure 62: Principal component 2 vs 1, representative burst sample spanning 5180 bursts. Data points clustered into one grouping.

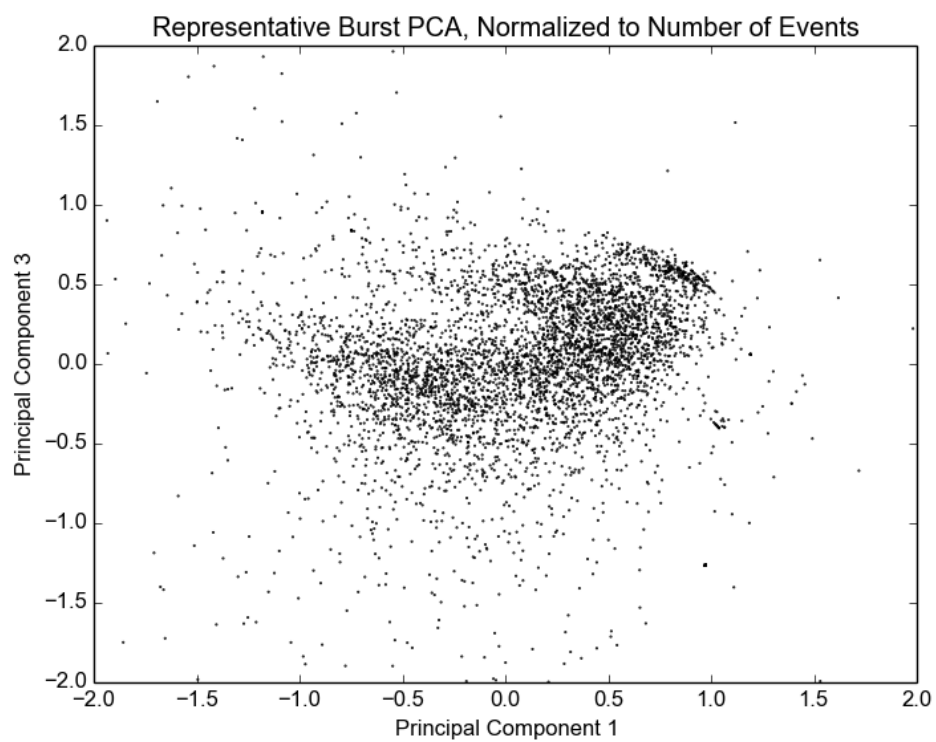


Figure 63: Principal component 3 vs 1, representative burst sample spanning 5180 bursts. Data points clustered into one grouping.

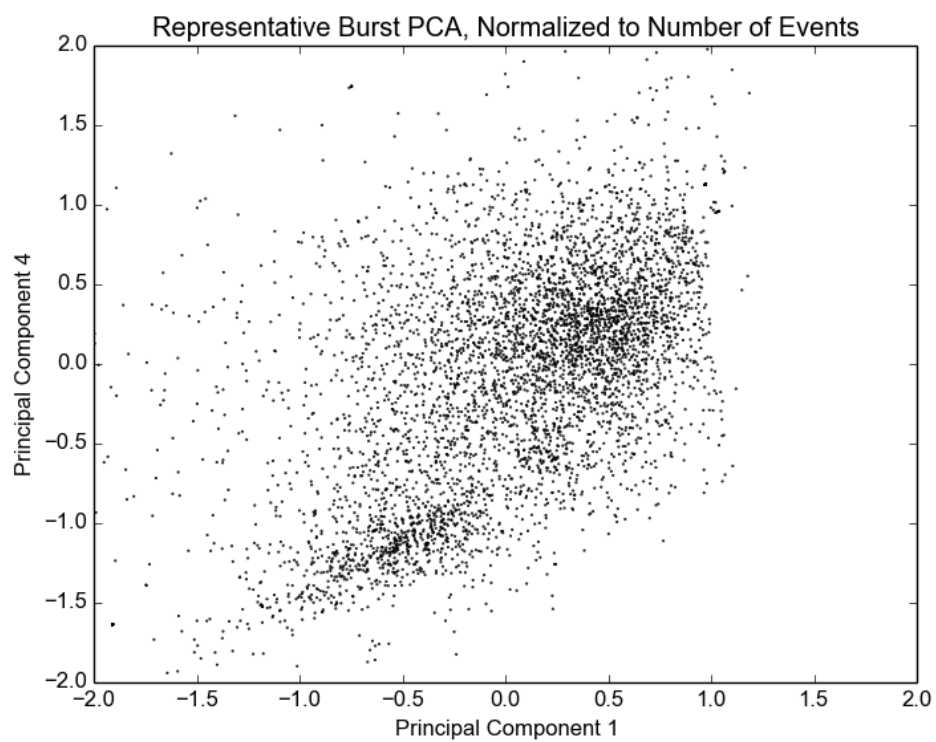


Figure 64: Principal component 4 vs 1, representative burst sample spanning 5180 bursts. Data points clustered into one grouping.

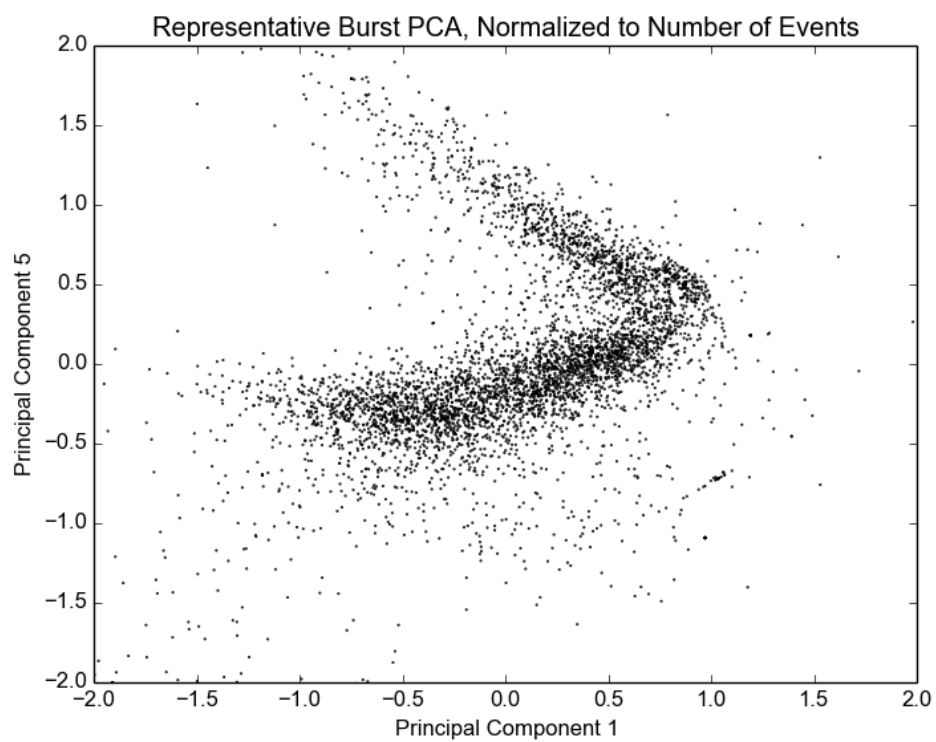


Figure 65: Principal component 5 vs 1, representative burst sample spanning 5180 bursts. 2 visible correlations of data points.

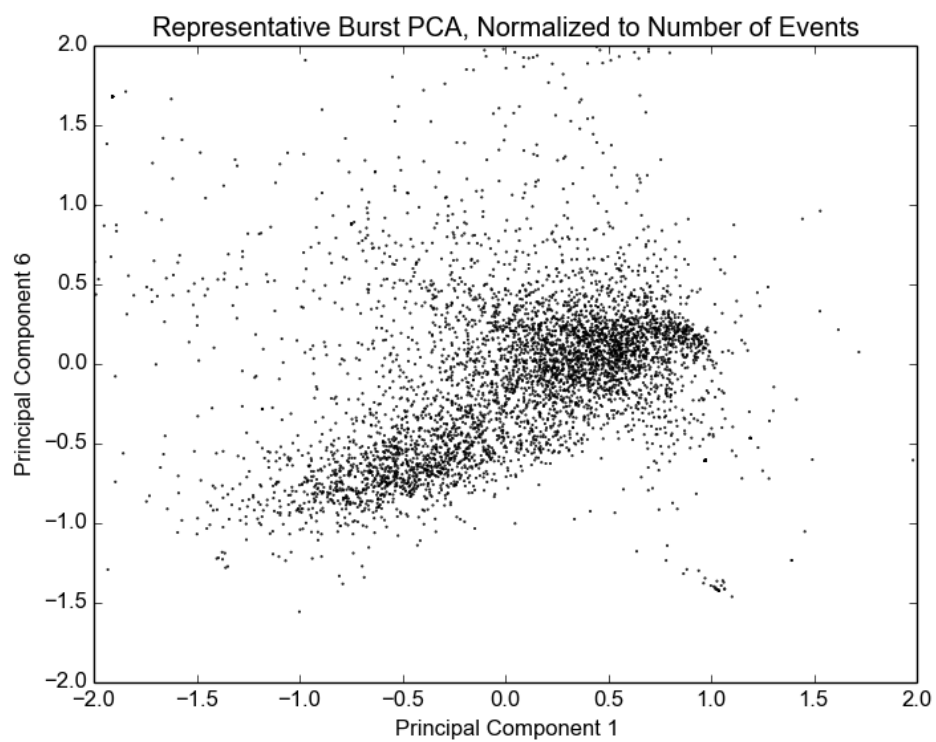


Figure 66: Principal component 6 vs 1, representative burst sample spanning 5180 bursts. 2 visible concentrations of data points.

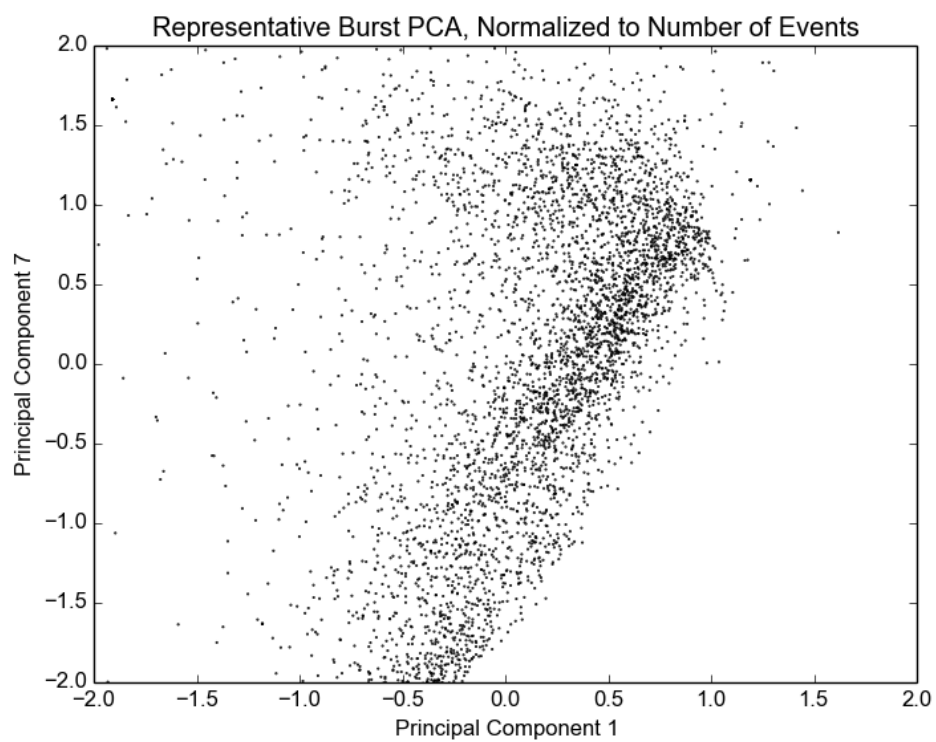


Figure 67: Principal component 7 vs 1, representative burst sample spanning 5180 bursts. Data points clustered into one grouping.

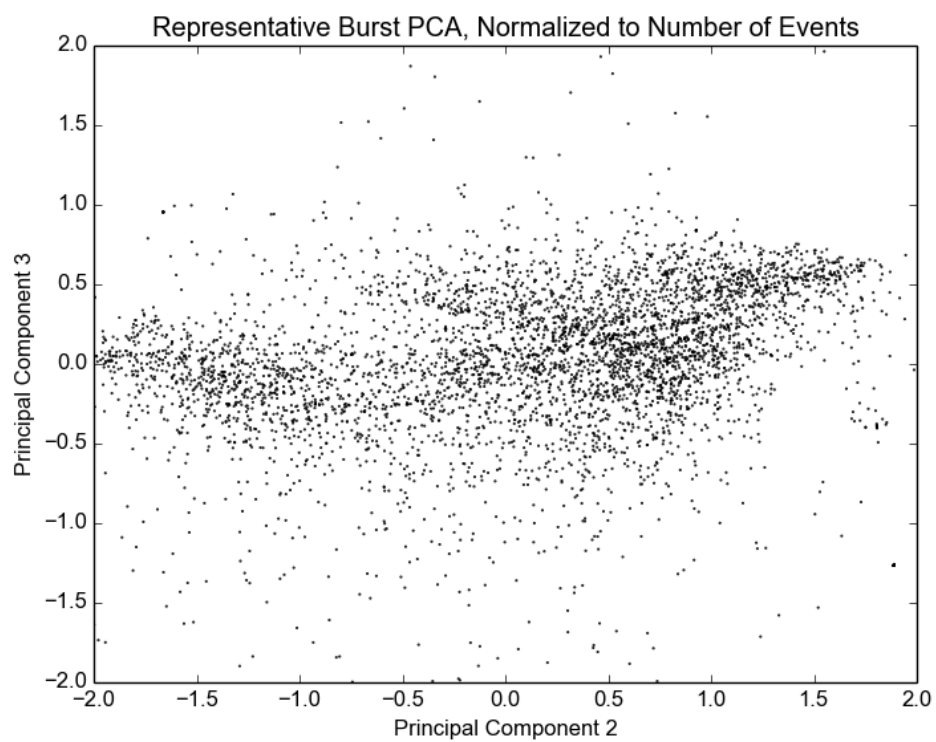


Figure 68: Principal component 3 vs 2, representative burst sample spanning 5180 bursts. Data points clustered into one grouping.

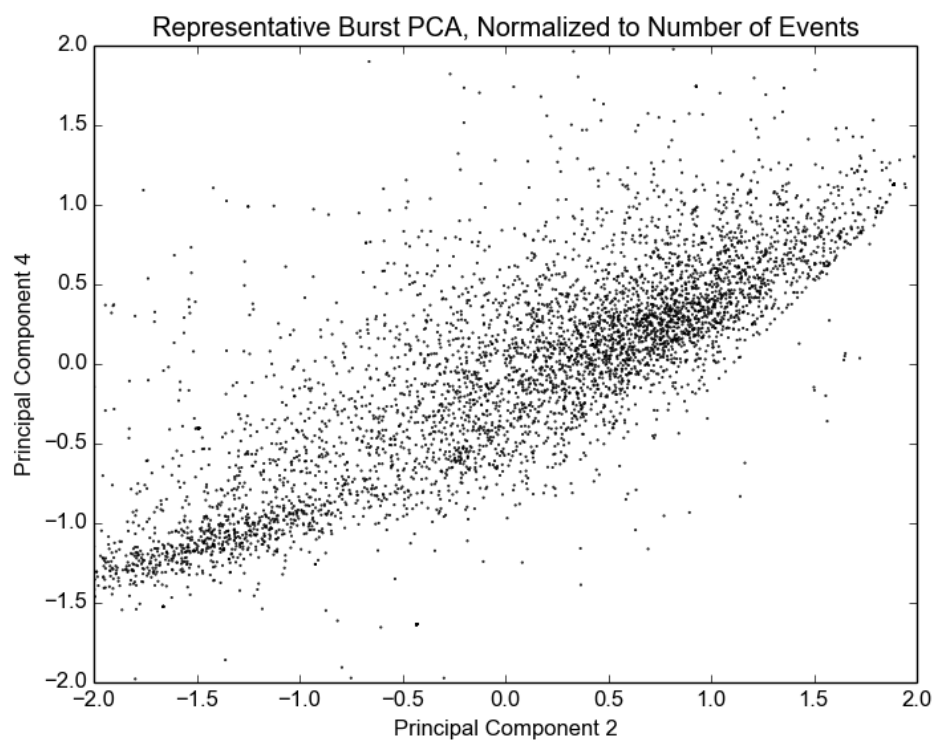


Figure 69: Principal component 4 vs 2, representative burst sample spanning 5180 bursts. Data points clustered into one grouping.

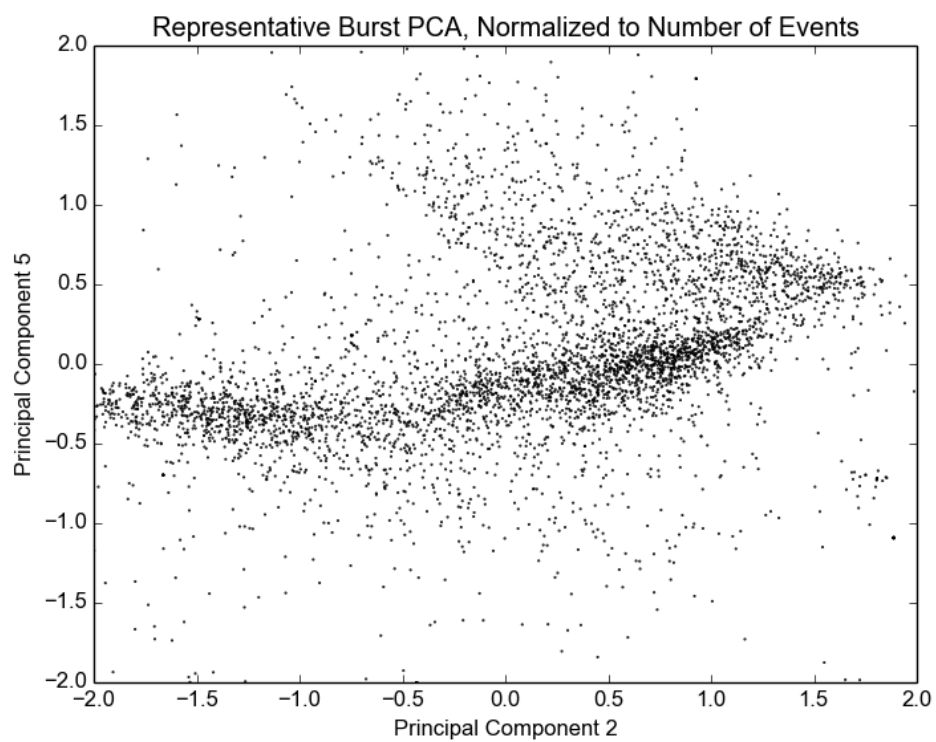


Figure 70: Principal component 5 vs 2, representative burst sample spanning 5180 bursts. 3 visible concentrations of data points.

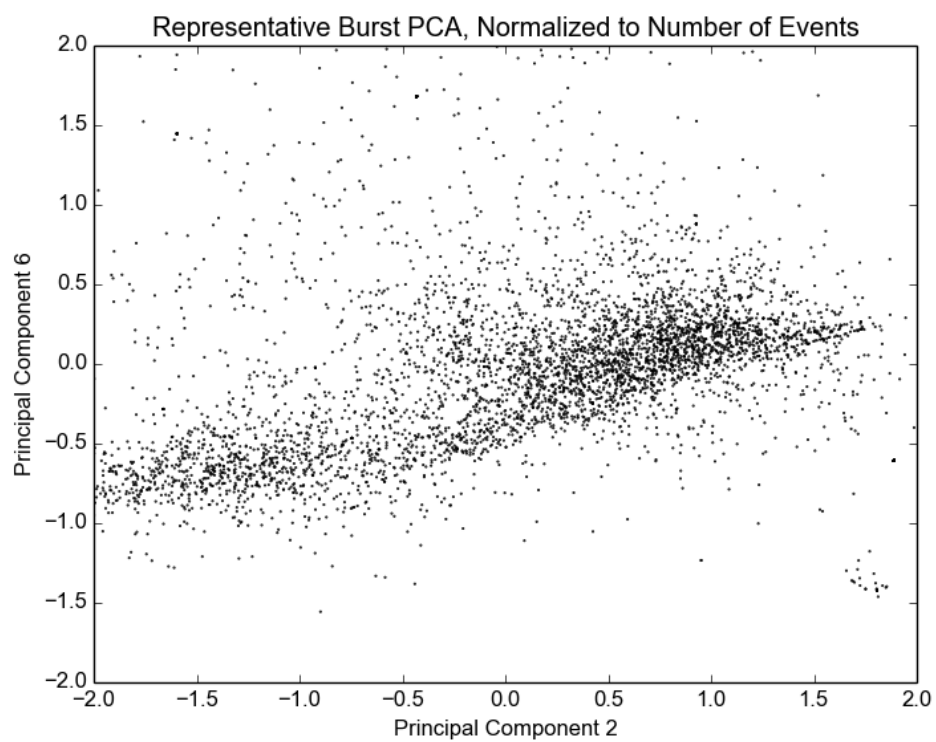


Figure 71: Principal component 6 vs 2, representative burst sample spanning 5180 bursts. 2 visible concentrations of data points.

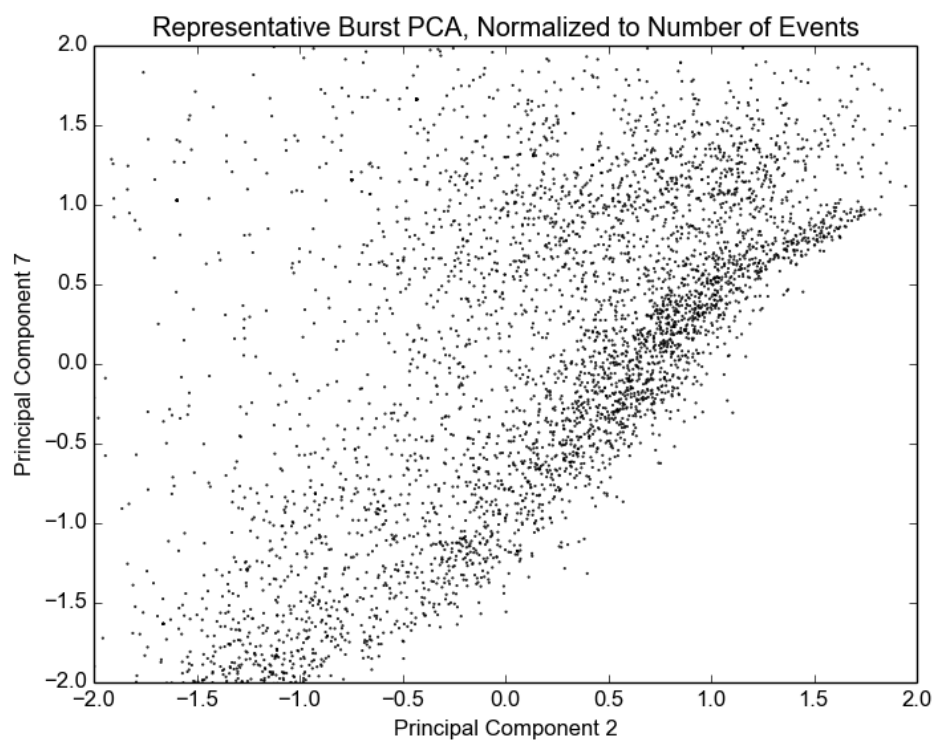


Figure 72: Principal component 7 vs 2, representative burst sample spanning 5180 bursts. Data points clustered into one grouping.

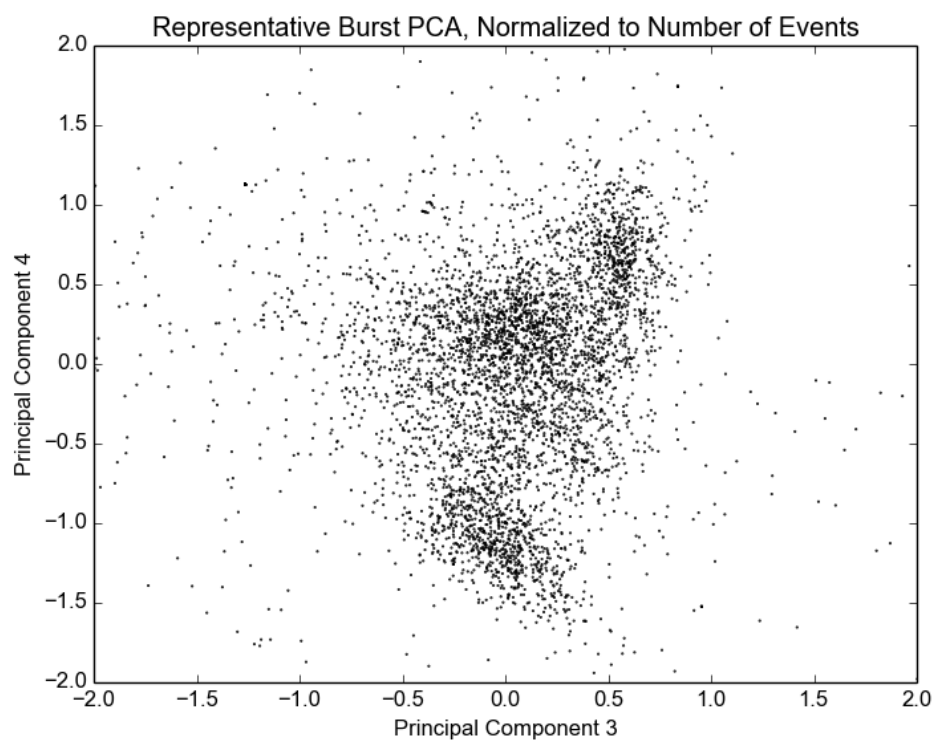


Figure 73: Principal component 4 vs 3, representative burst sample spanning 5180 bursts. 3 visible concentrations of data points.

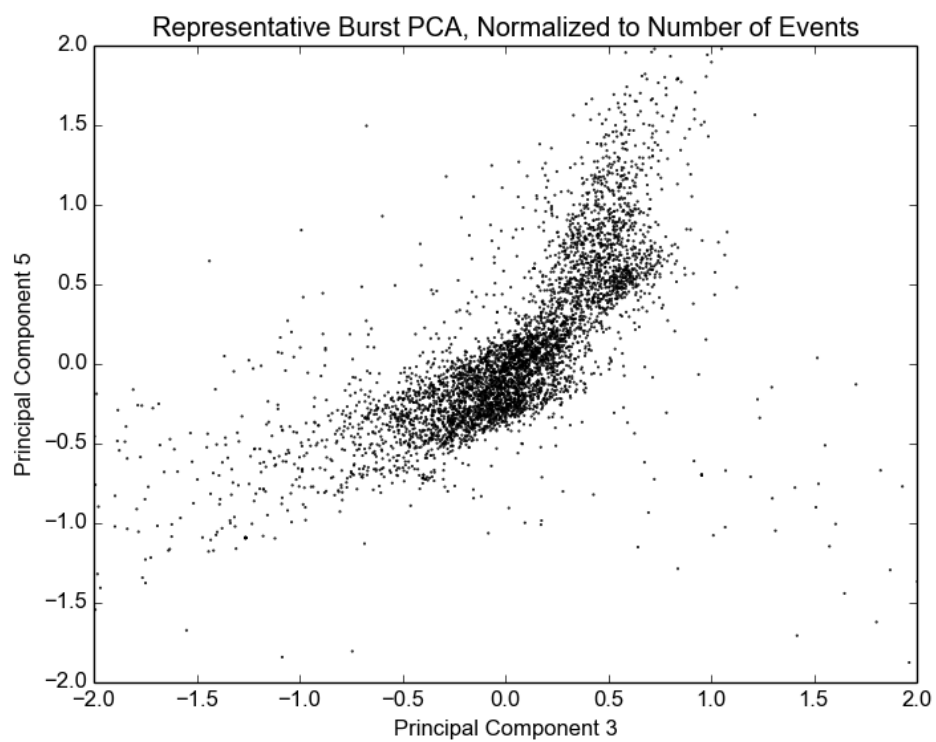


Figure 74: Principal component 5 vs 3, representative burst sample spanning 5180 bursts. 2 visible concentrations of data points. Notably more densely populated than other PCA plots.

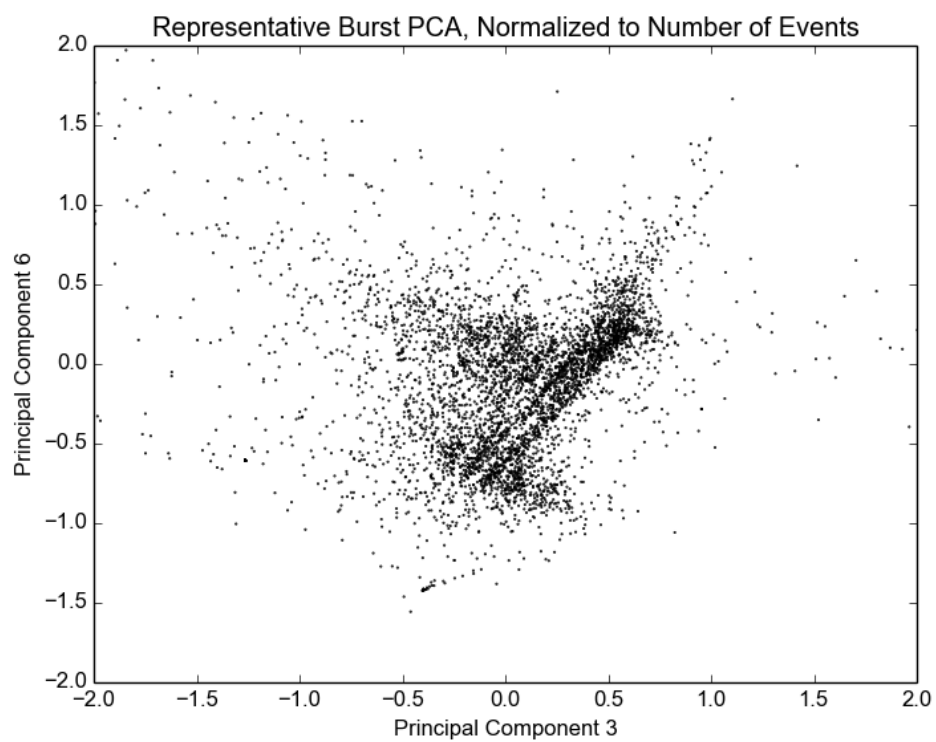


Figure 75: Principal component 6 vs 3, representative burst sample spanning 5180 bursts. A "ripple" effect can be seen along three higher concentrations of data points.

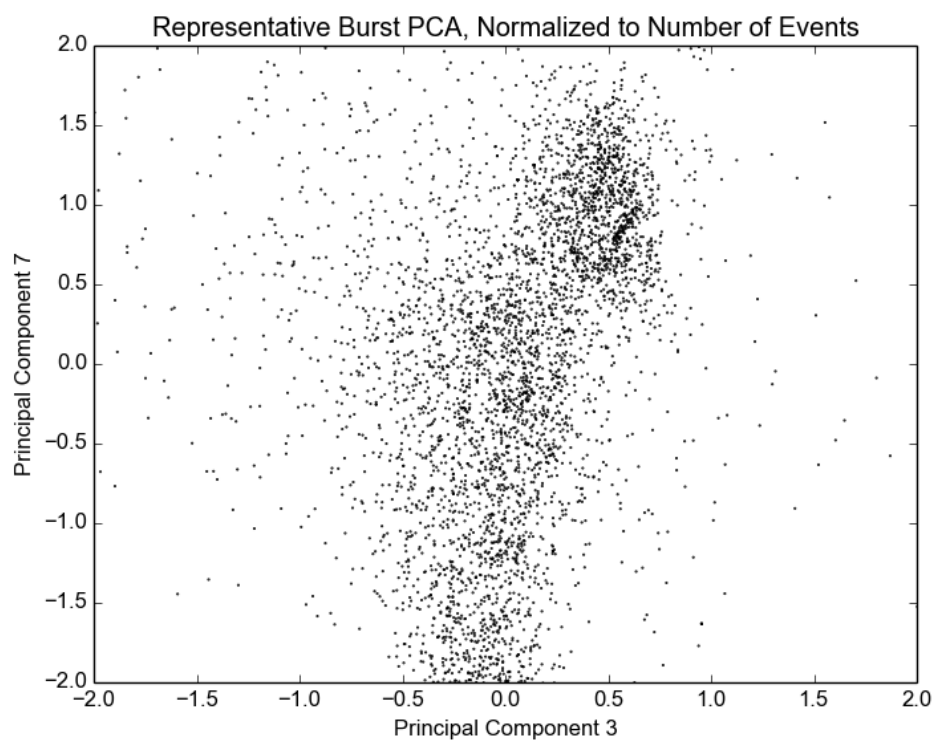


Figure 76: Principal component 7 vs 3, representative burst sample spanning 5180 bursts. Data points are spread evenly through one grouping.

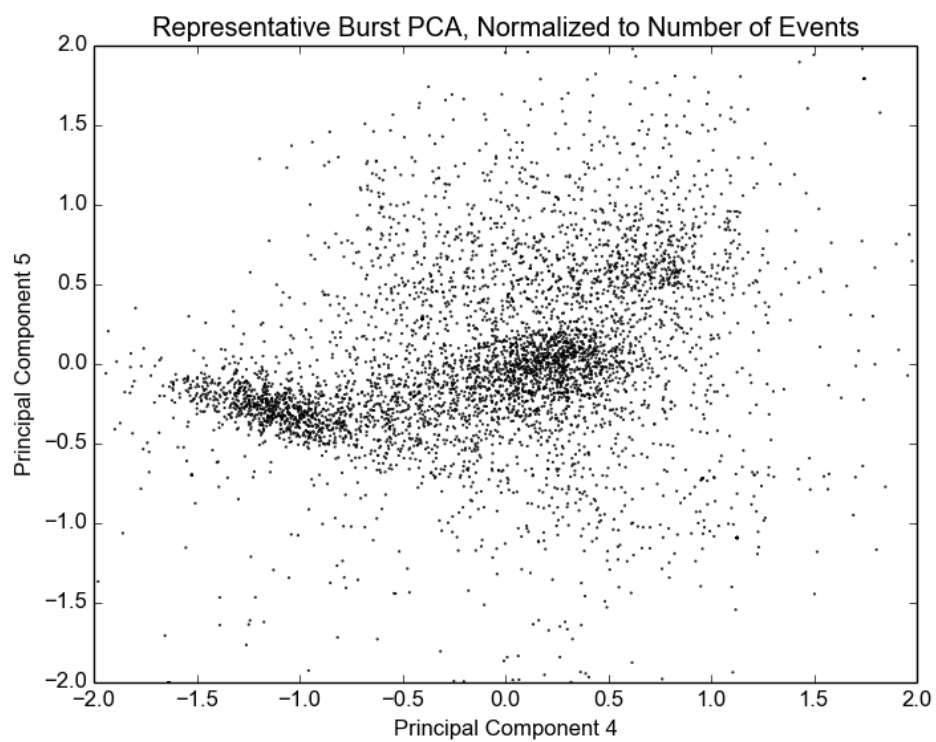


Figure 77: Principal component 5 vs 4, representative burst sample spanning 5180 bursts. 3 visible concentrations of data points.

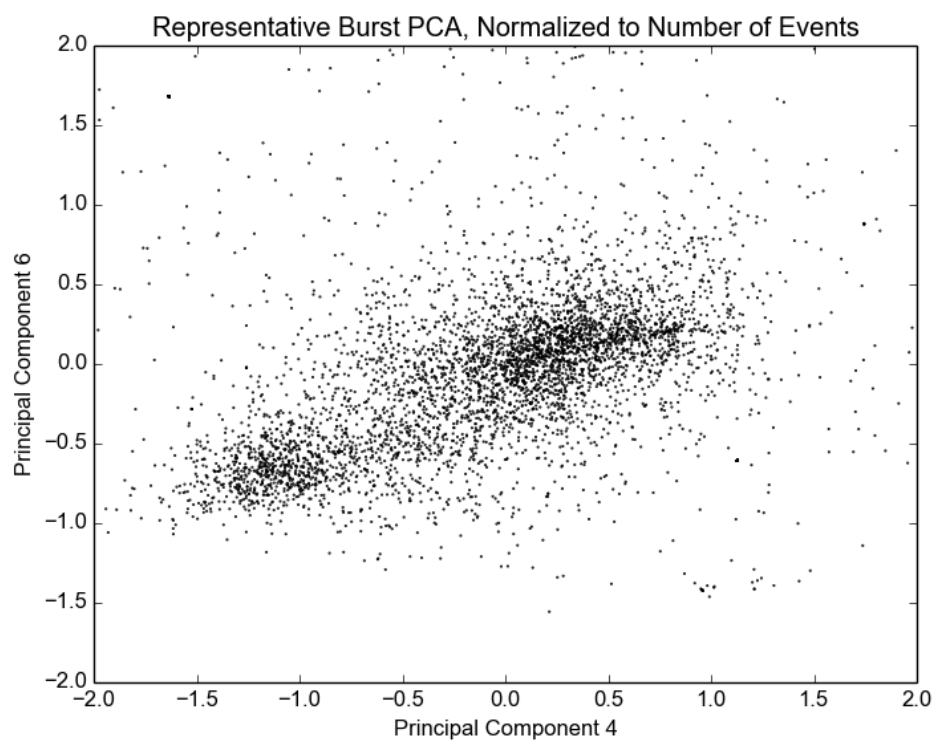


Figure 78: Principal component 6 vs 4, representative burst sample spanning 5180 bursts. 2 visible concentrations of data points.

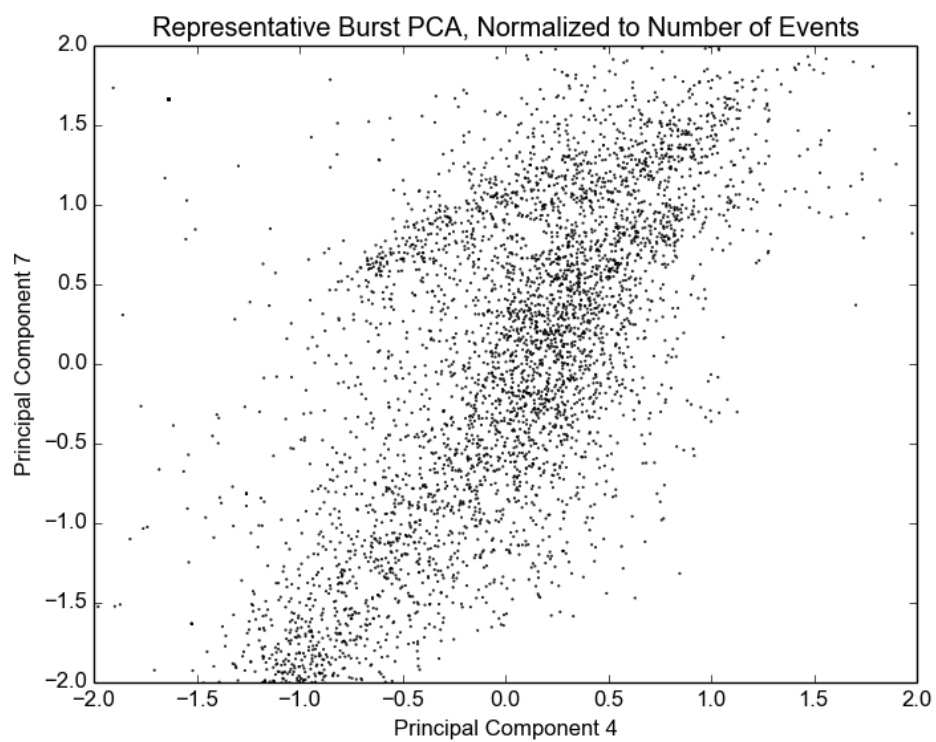


Figure 79: Principal component 7 vs 4, representative burst sample spanning 5180 bursts. Data points are spread evenly through one large grouping.

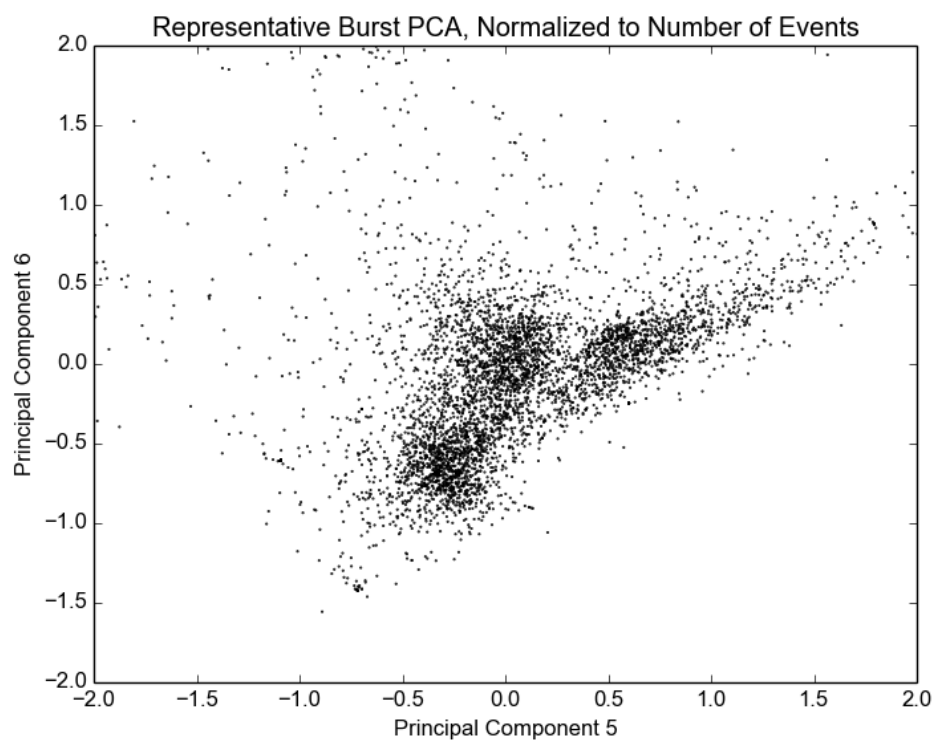


Figure 80: Principal component 6 vs 5, representative burst sample spanning 5180 bursts. 3 visible concentrations of data points. A "ripple" effect can be seen.

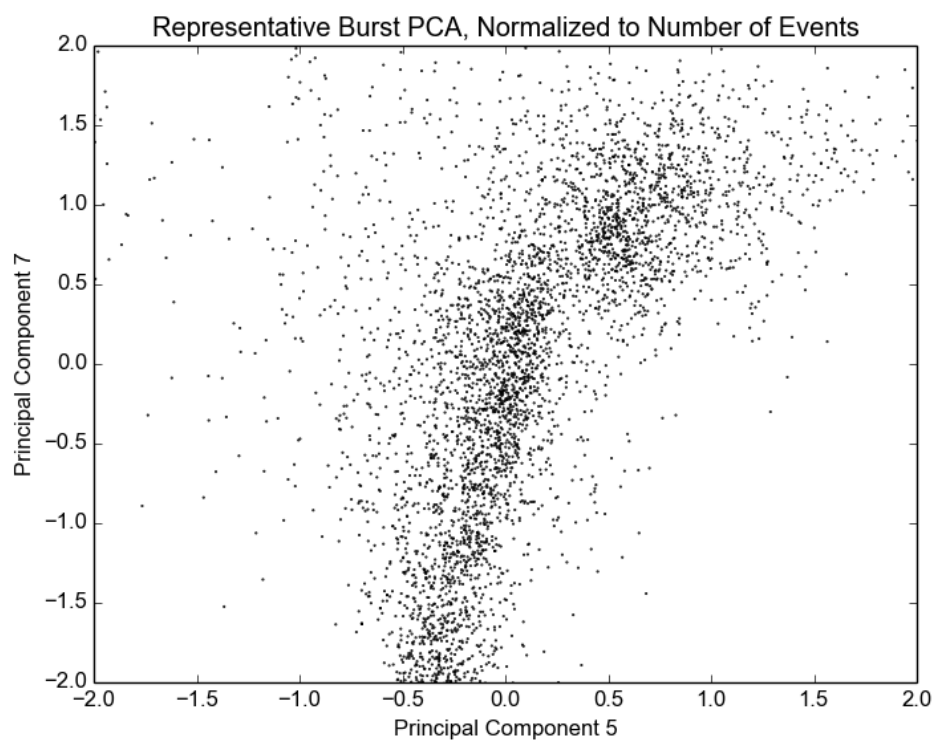


Figure 81: Principal component 7 vs 5, representative burst sample spanning 5180 bursts. Data points are spread evenly through one large grouping.

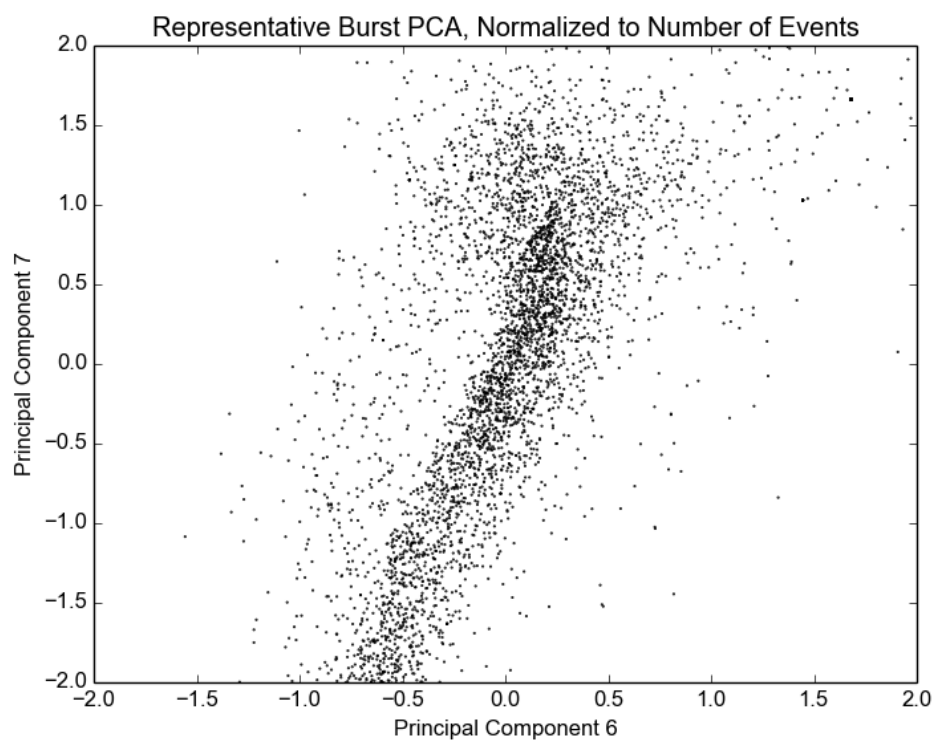


Figure 82: Principal component 7 vs 6, representative burst sample spanning 5180 bursts. Data points are spread evenly among one large grouping.

The first PCA plot in this section spans the greatest total variance of all selections and is the most important to PCA. 3 main groupings can be seen qualitatively in the plot with a region of lower density in the right half of the plot. This plot will be discussed in greater detail in chapter 7.

PCA plots which also show 3 primary groupings are the plots of principal components:

- 2 vs 0
- 3 vs 0
- 4 vs 0
- 6 vs 0
- 3 vs 1
- 6 vs 1
- 5 vs 2
- 4 vs 3
- 6 vs 3
- 5 vs 4
- 6 vs 5

PCA plots which are indicative of structure and correlation are the plots of principal components:

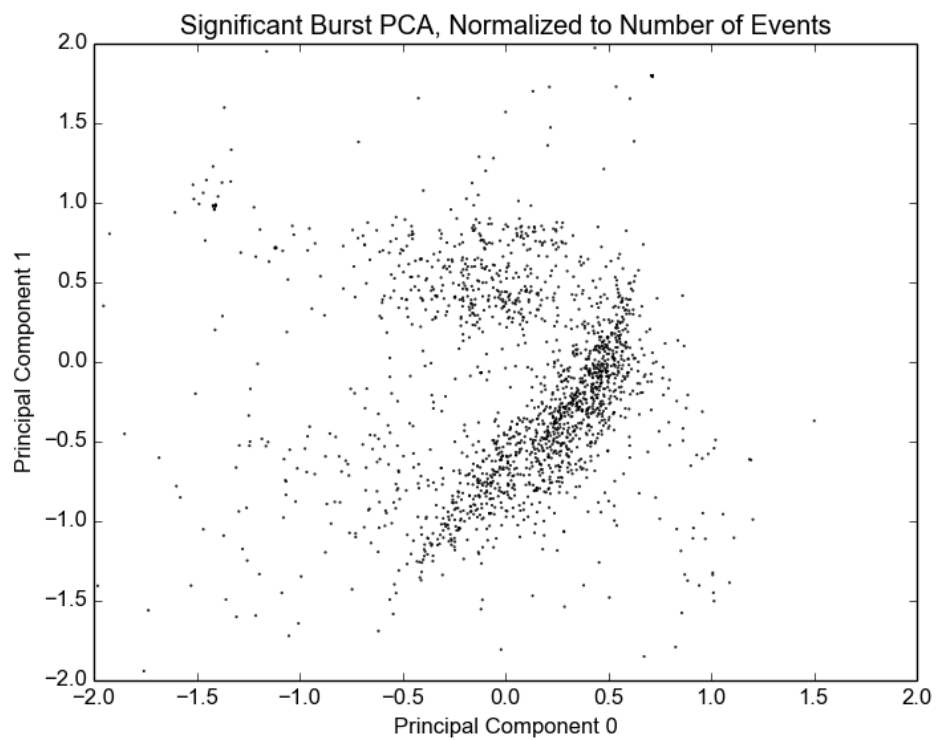
- 5 vs 0
- 7 vs 0
- 5 vs 1
- 5 vs 3
- 6 vs 4

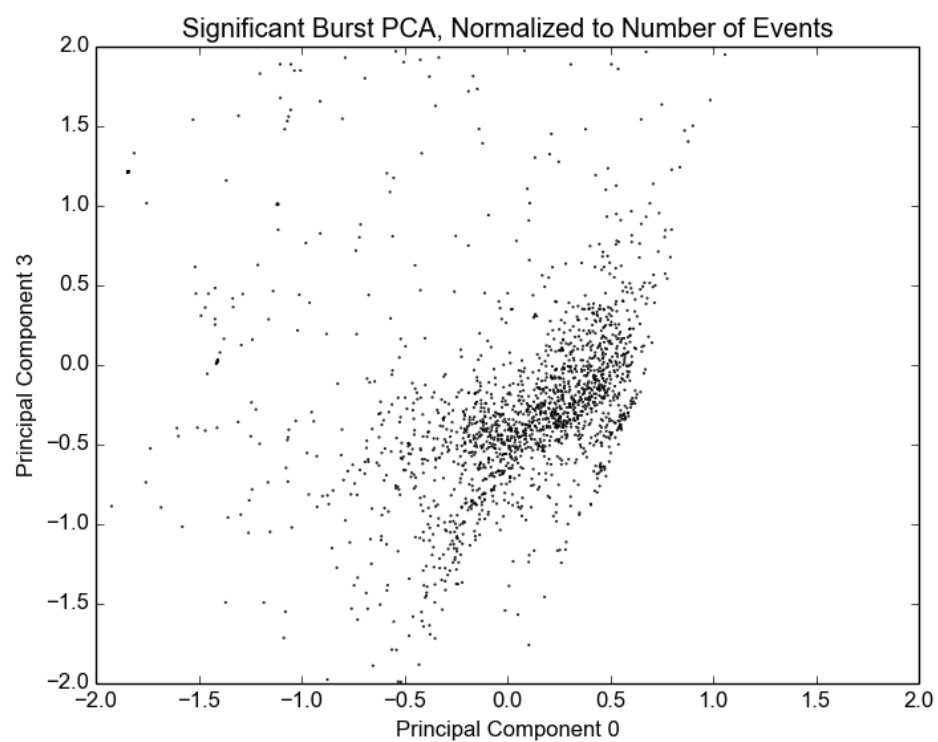
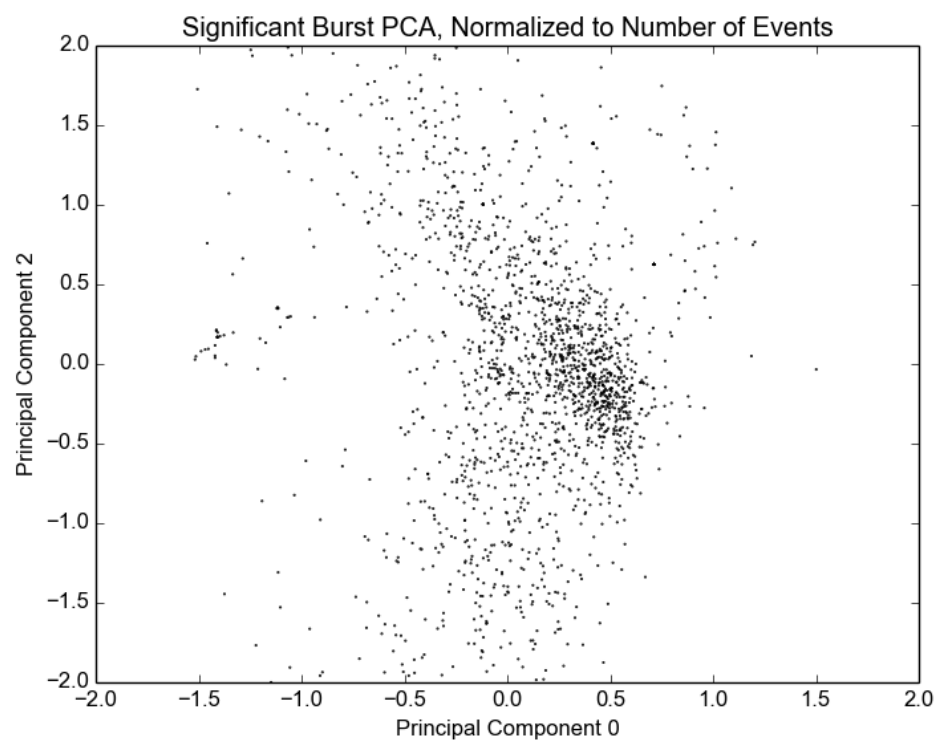
PCA plots which show no notable pattern are the plots of principal components:

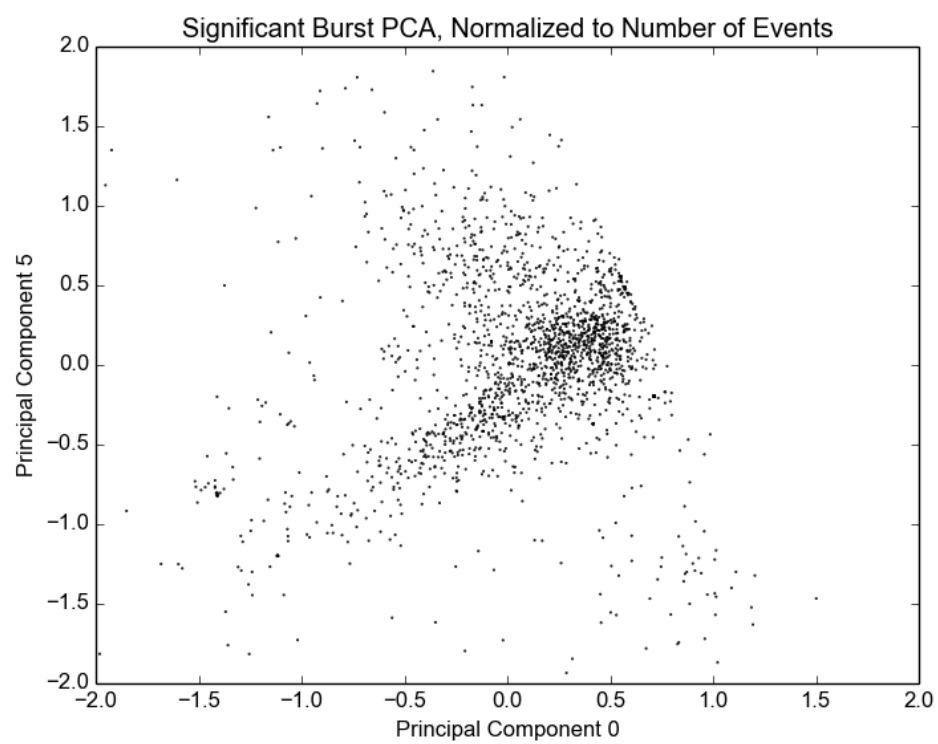
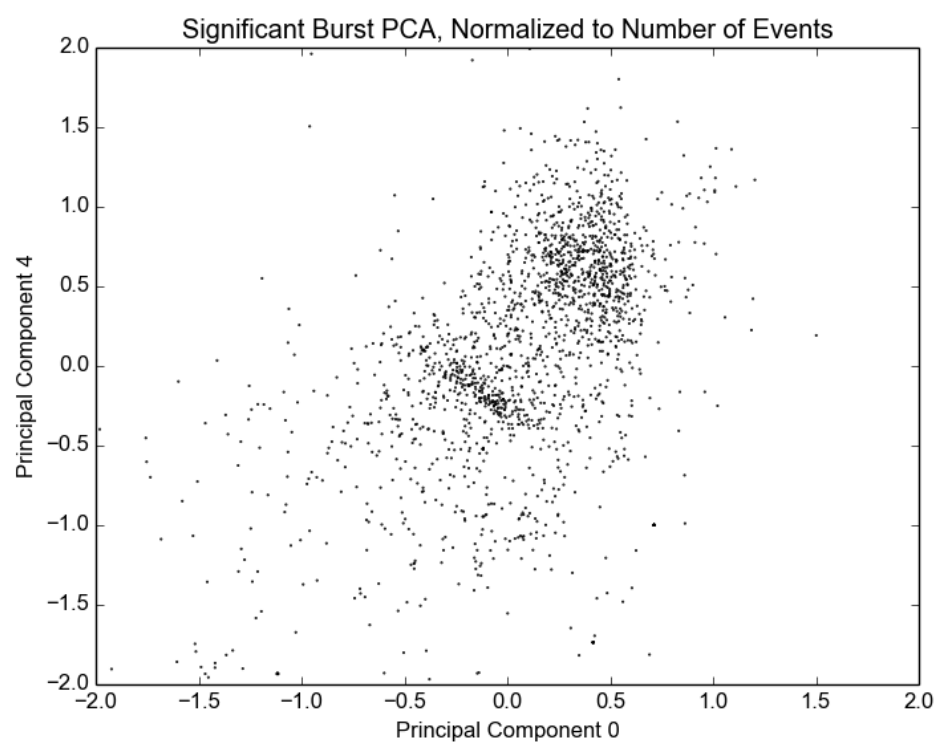
- 2 vs 1
- 4 vs 1
- 7 vs 1
- 3 vs 2
- 4 vs 2
- 6 vs 2
- 7 vs 2
- 7 vs 3
- 7 vs 4
- 7 vs 5
- 7 vs 6

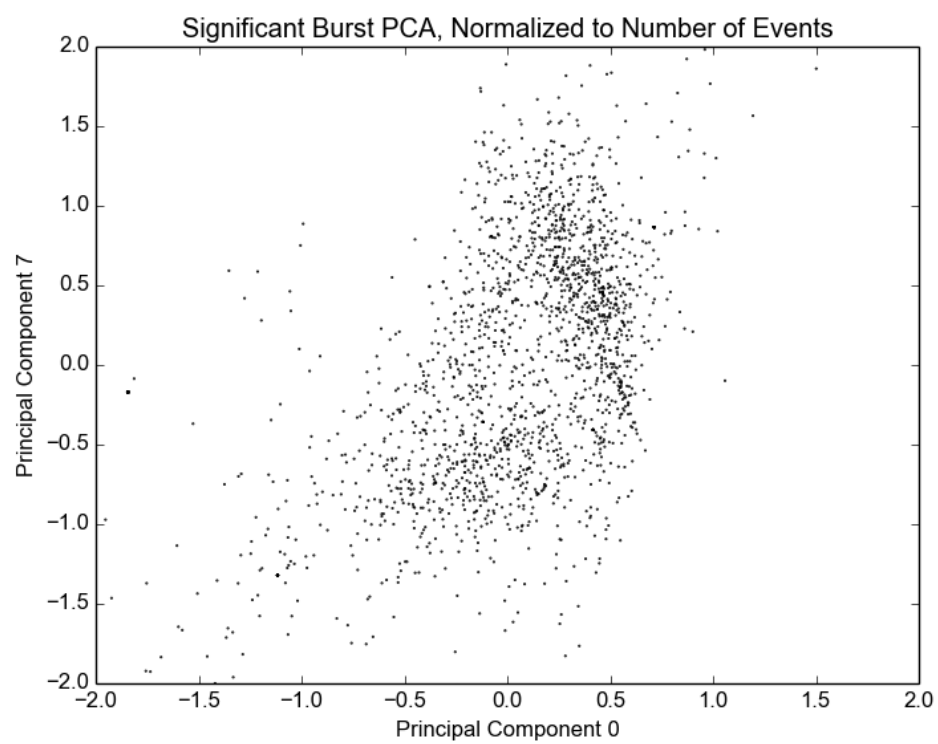
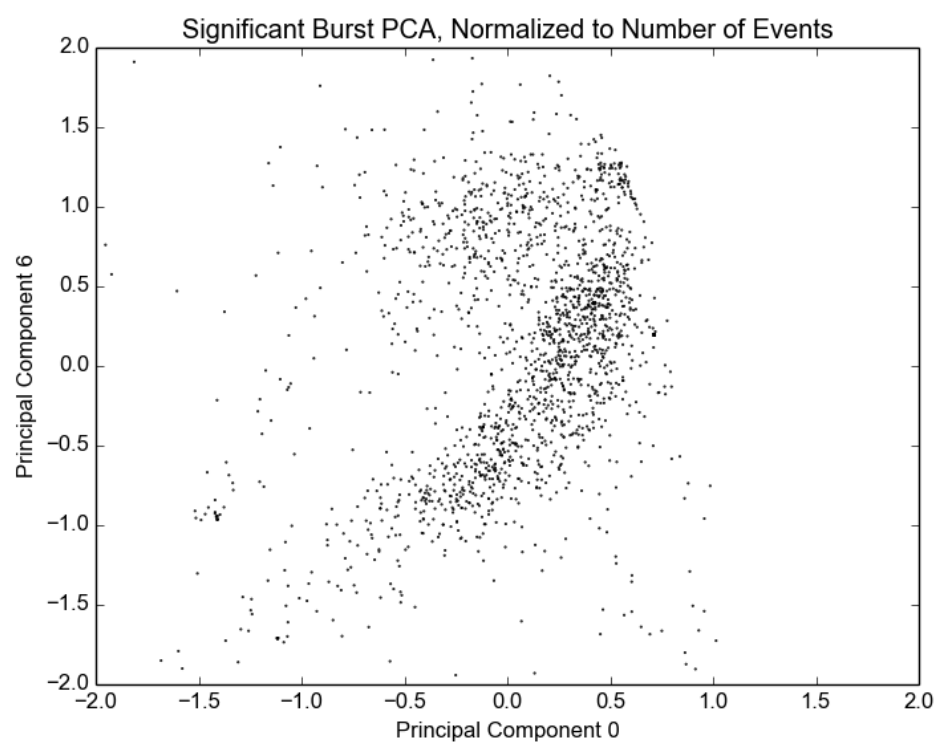
Note that higher order principal components tend to appear more in the lists which show some or no structure or correlation as opposed to lower order principal components tend show structure in the data. This is the effect of the procedural method through which the principal components are determined.

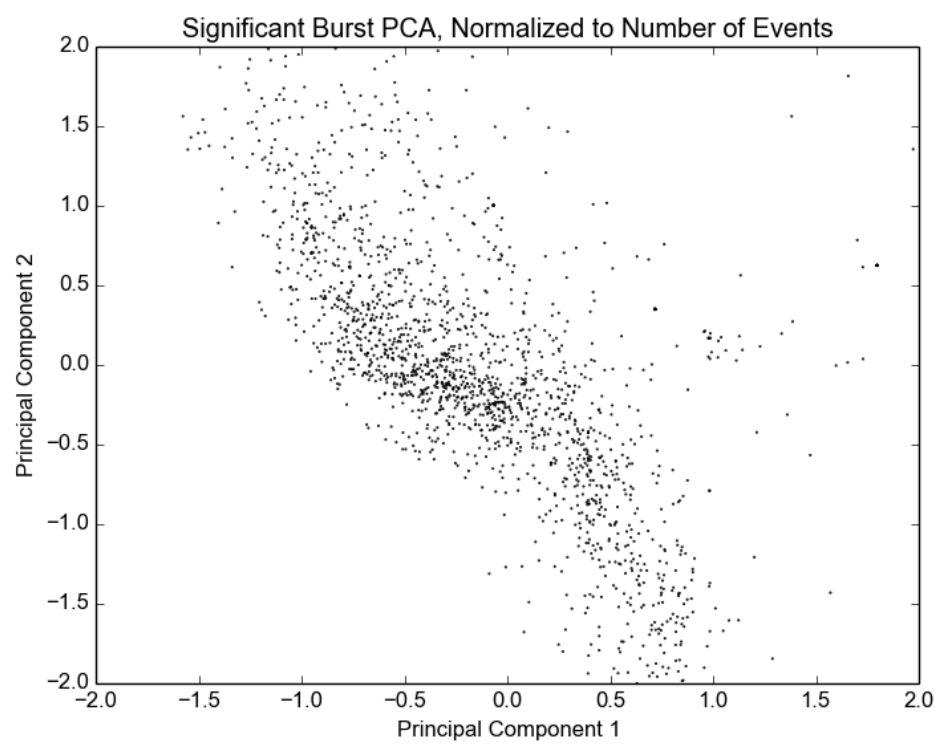
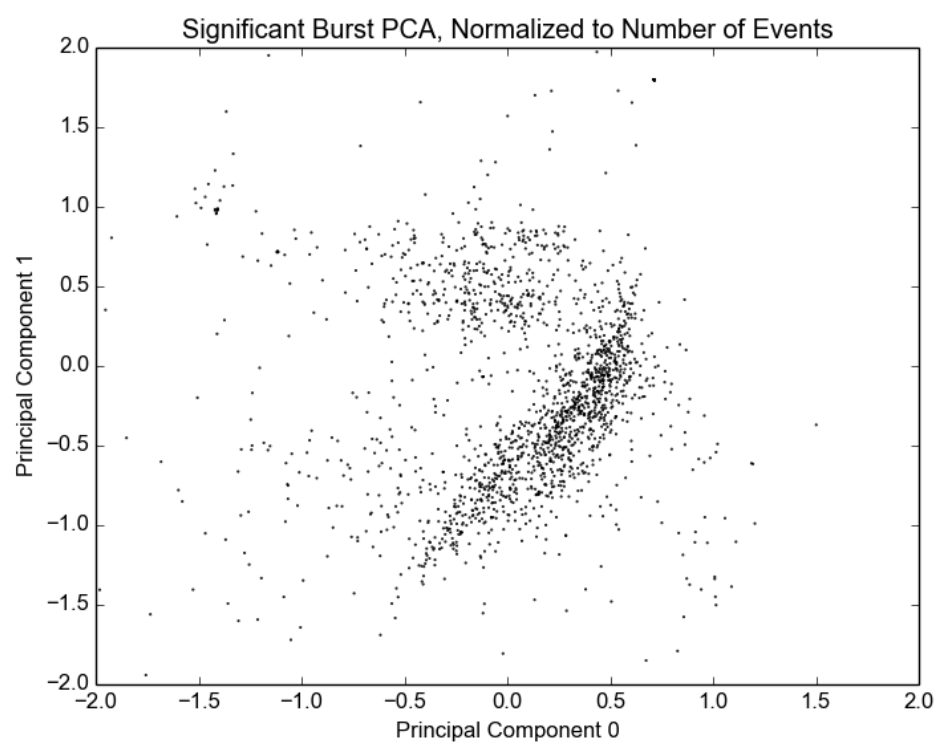
C.2 Significant Bursts PCA plots

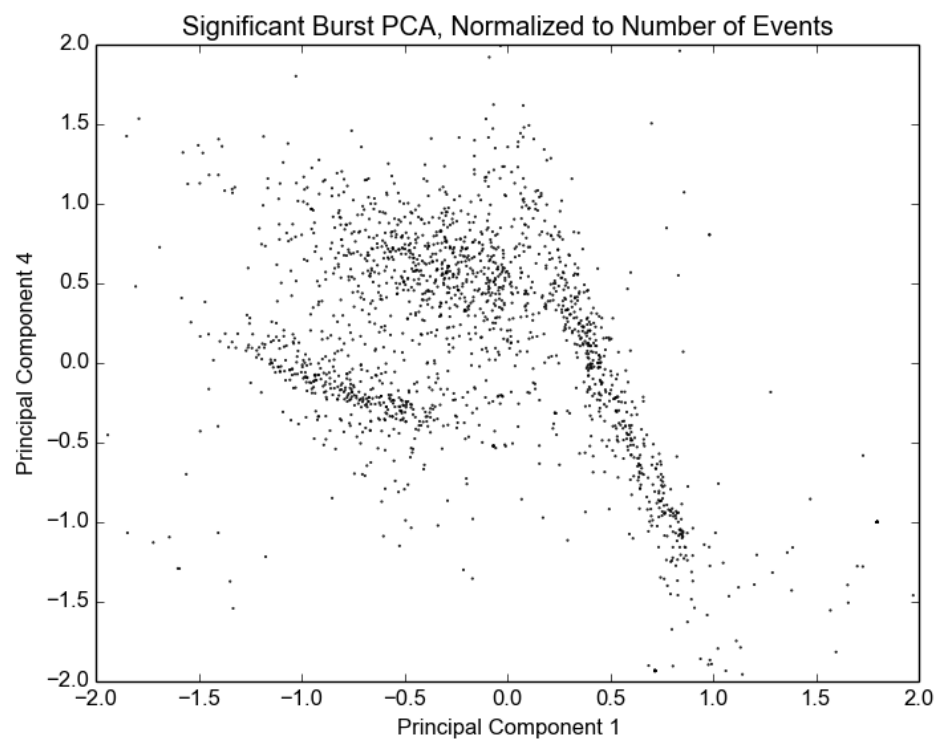
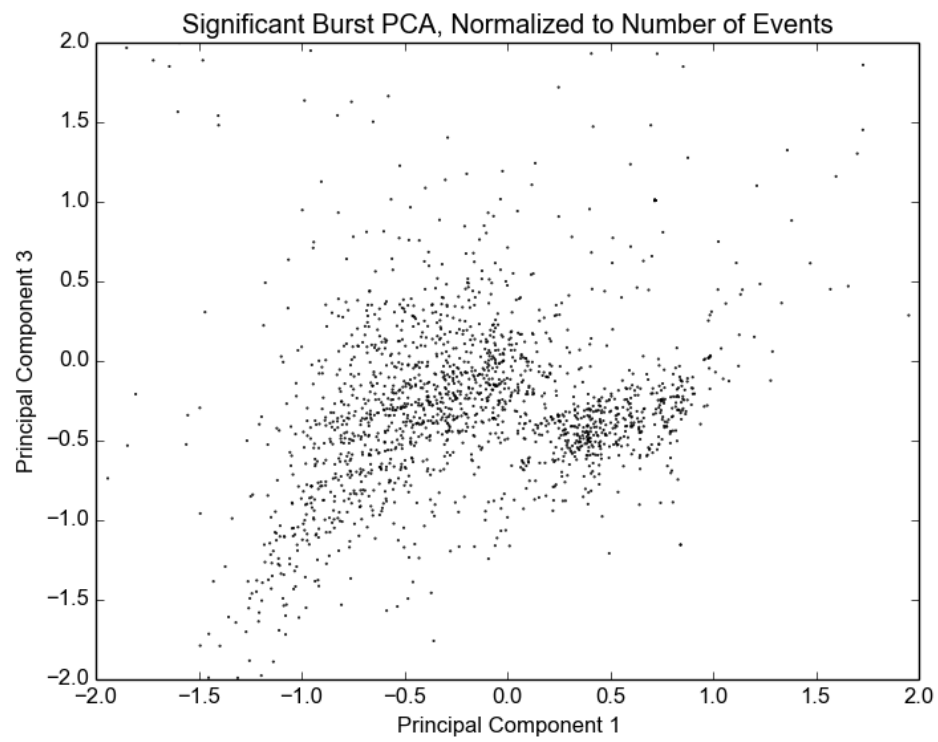


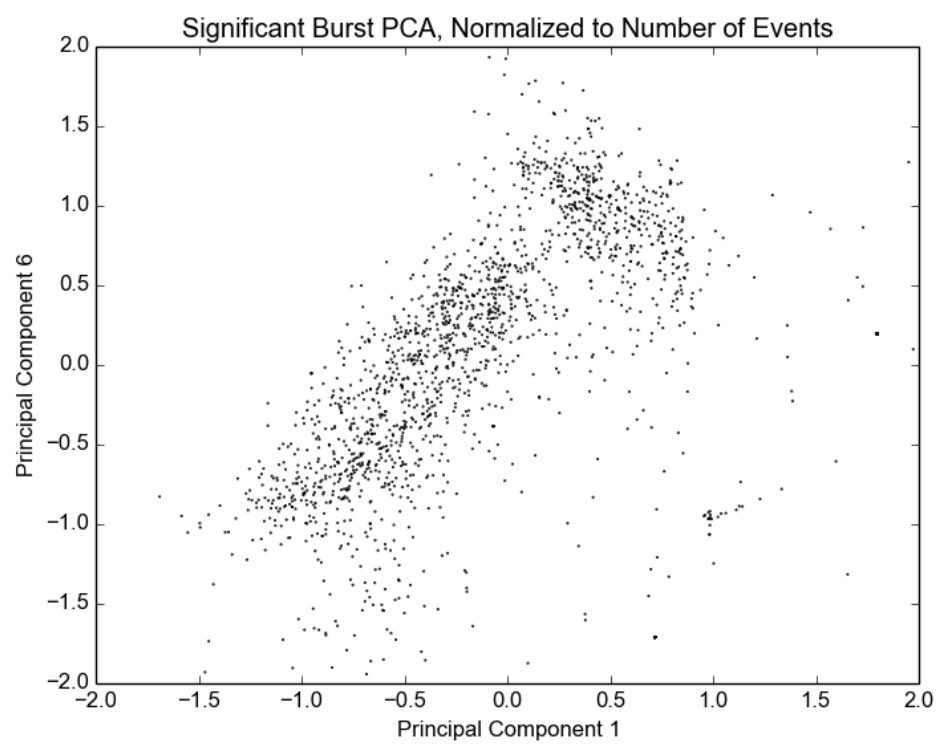
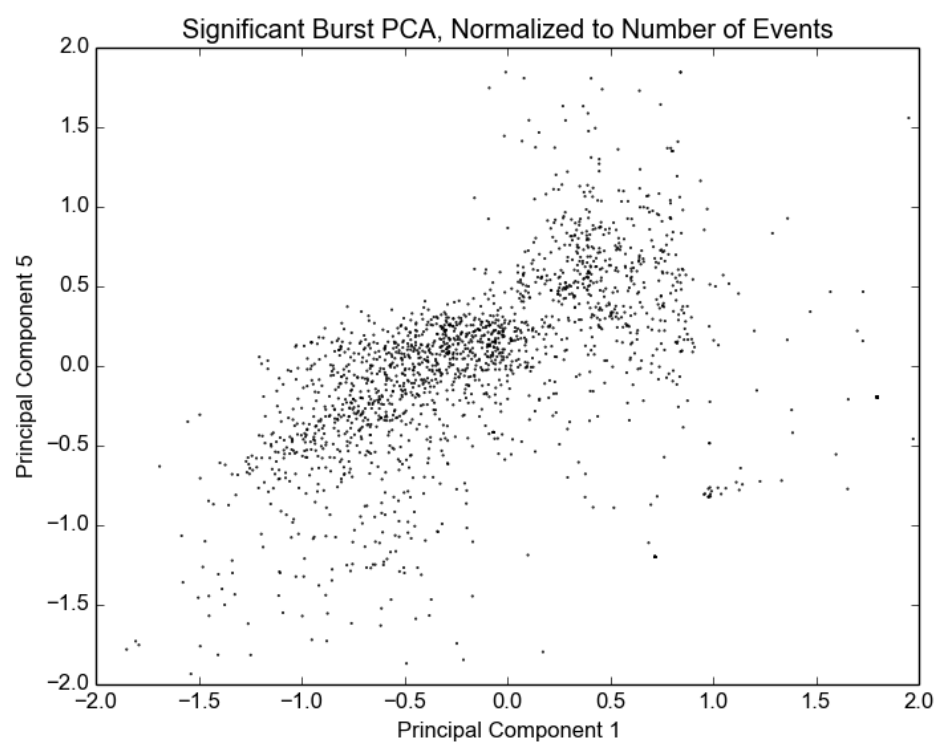


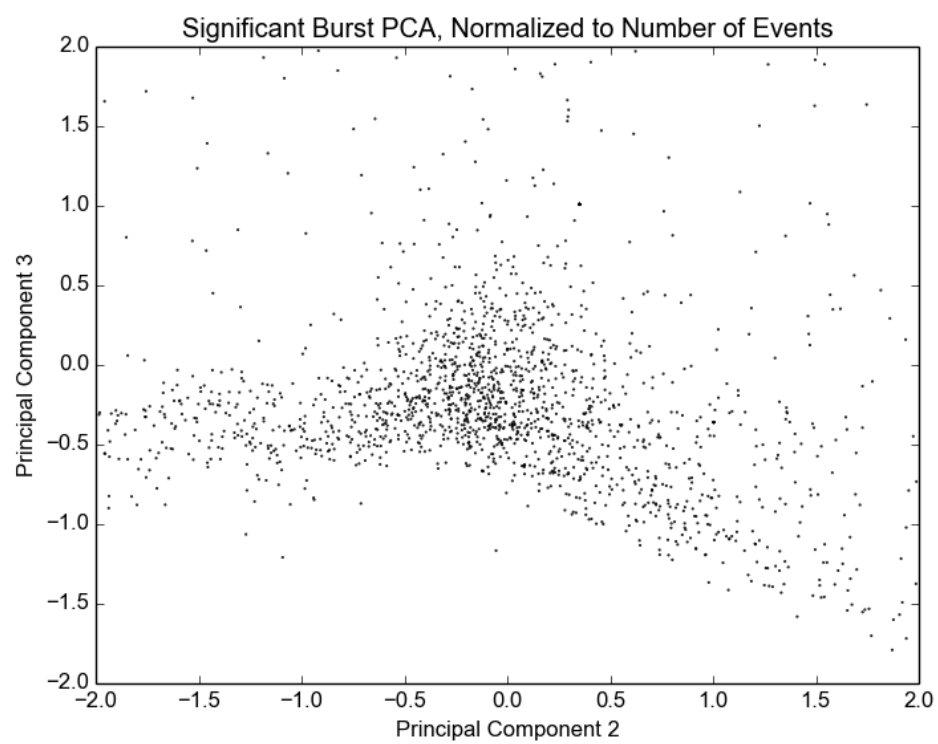
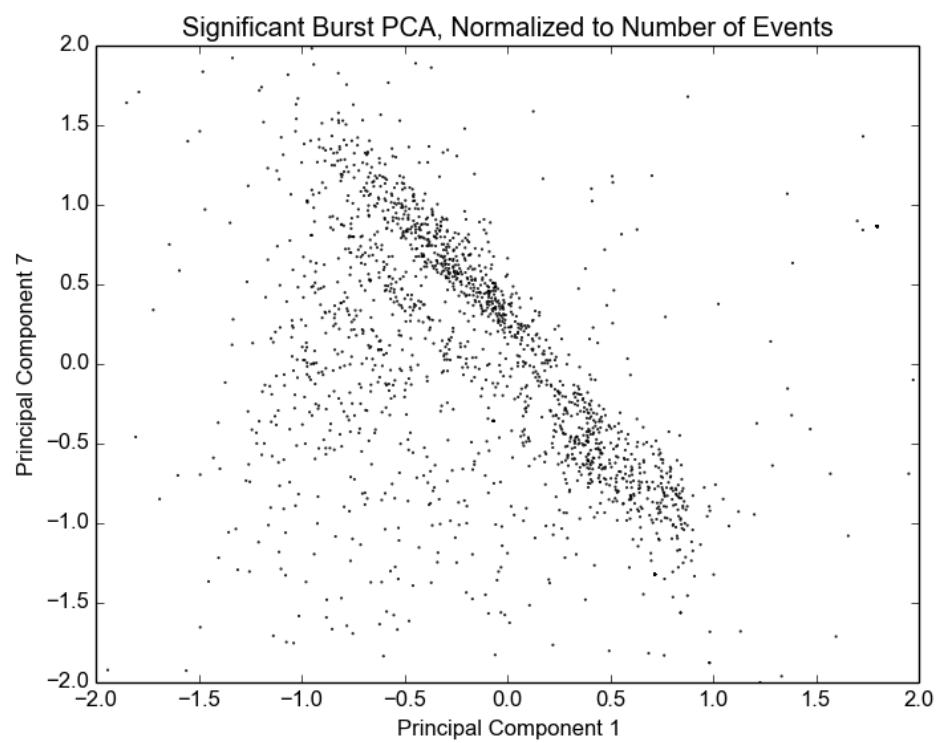


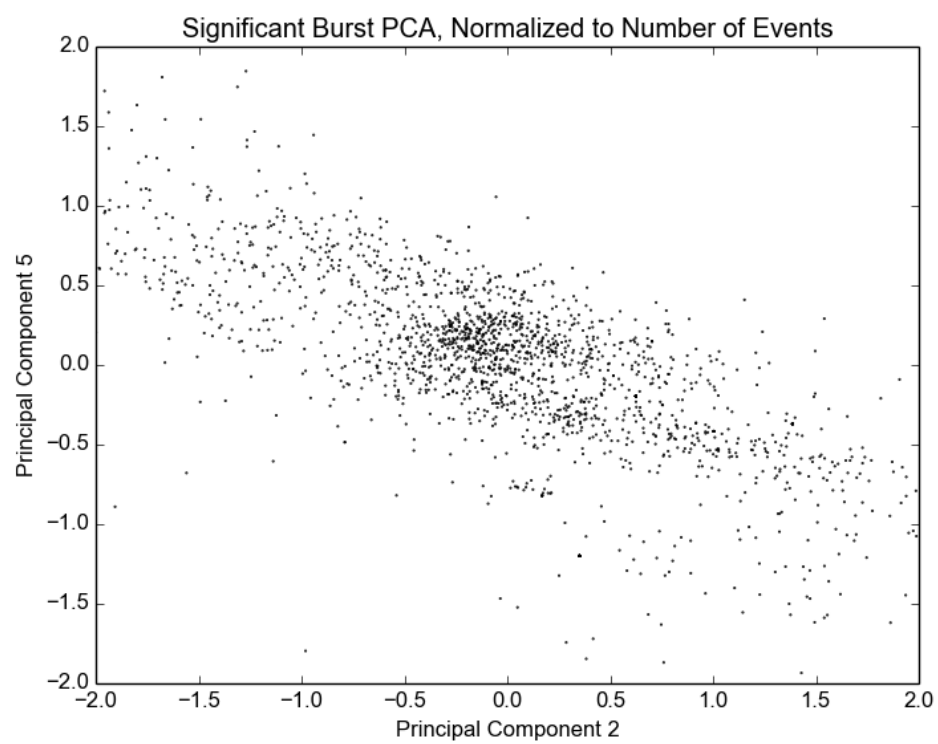
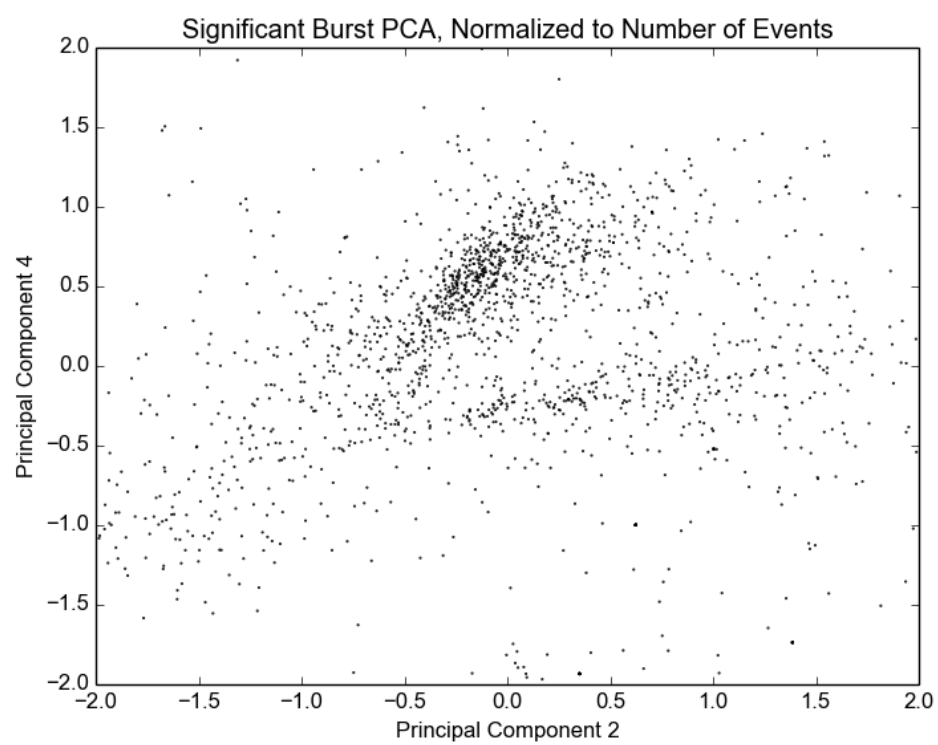


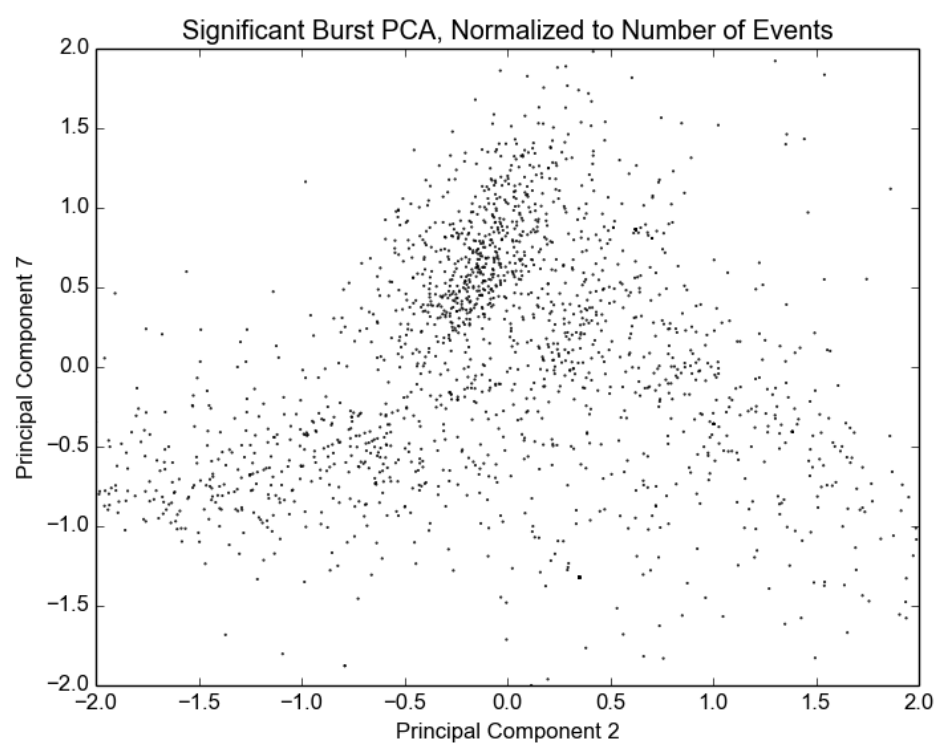
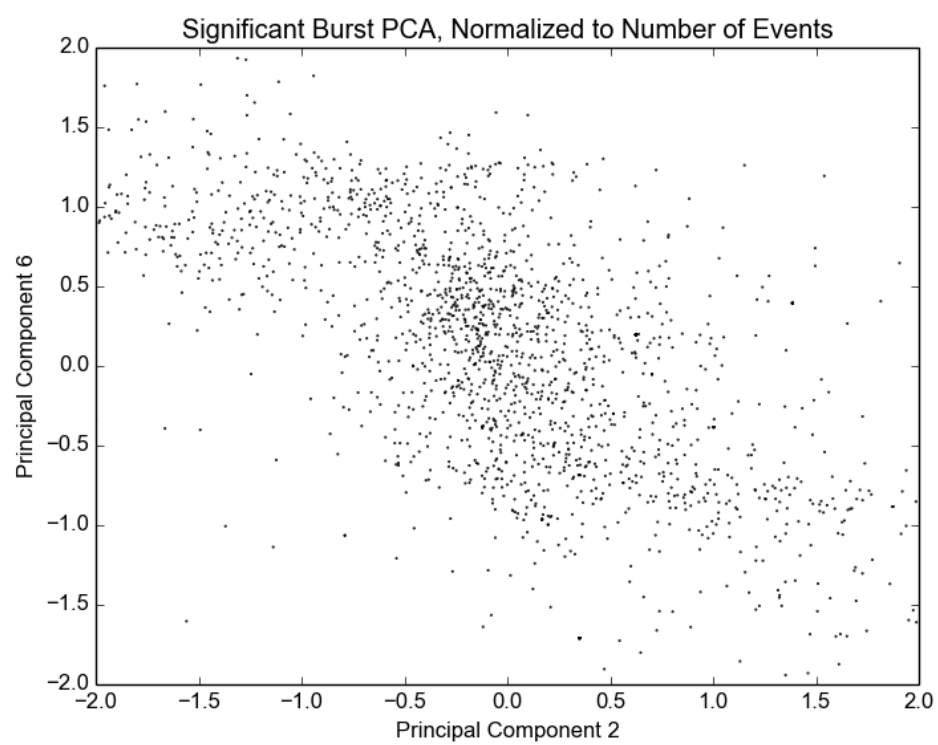


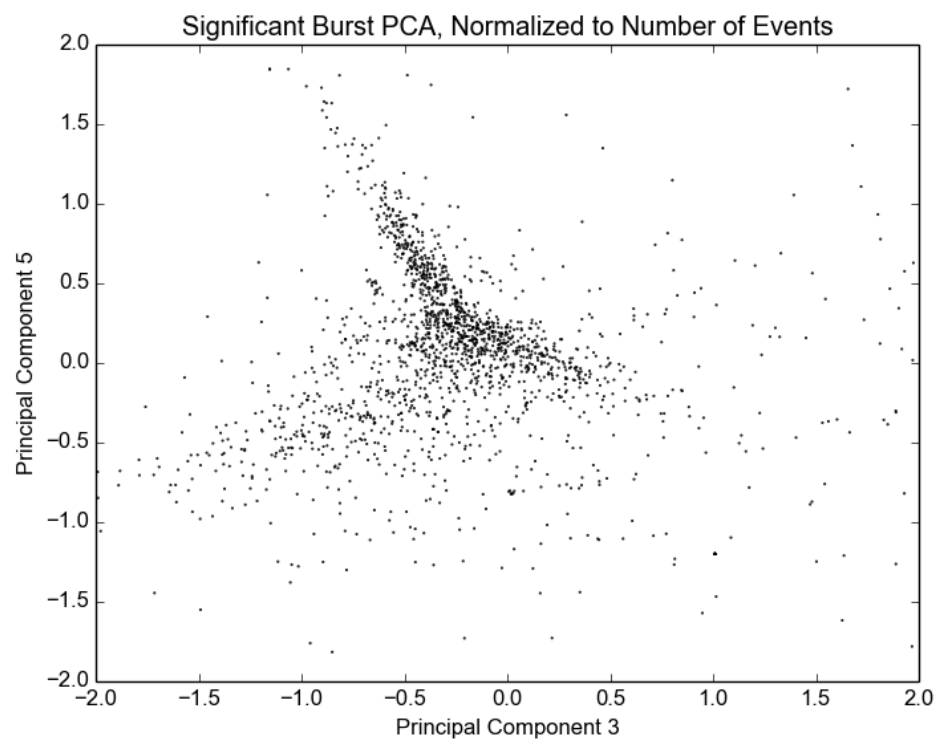
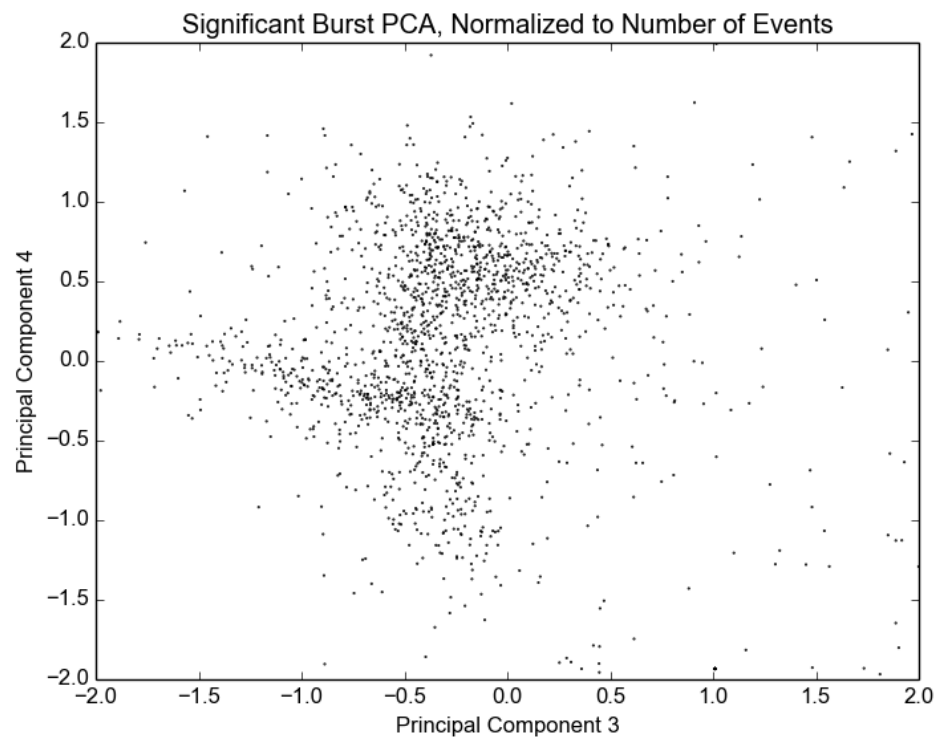


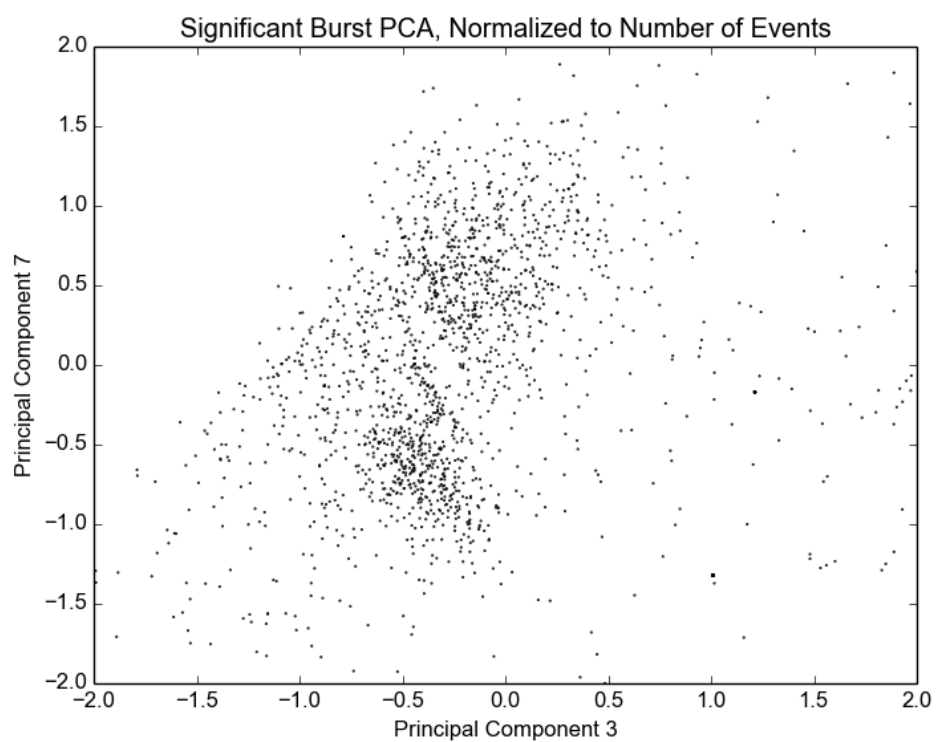
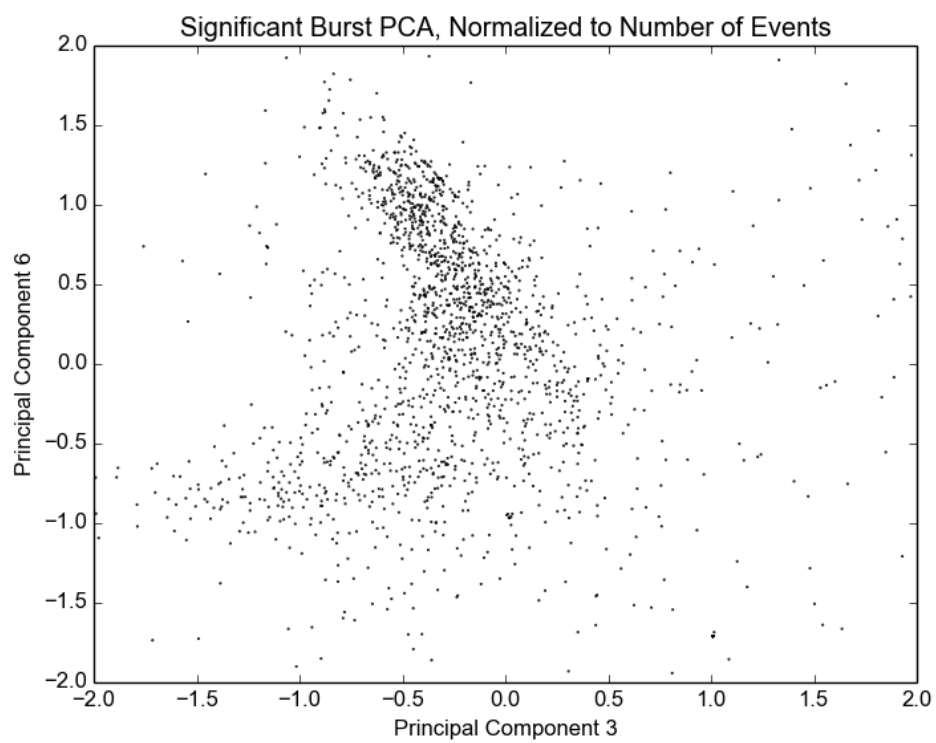


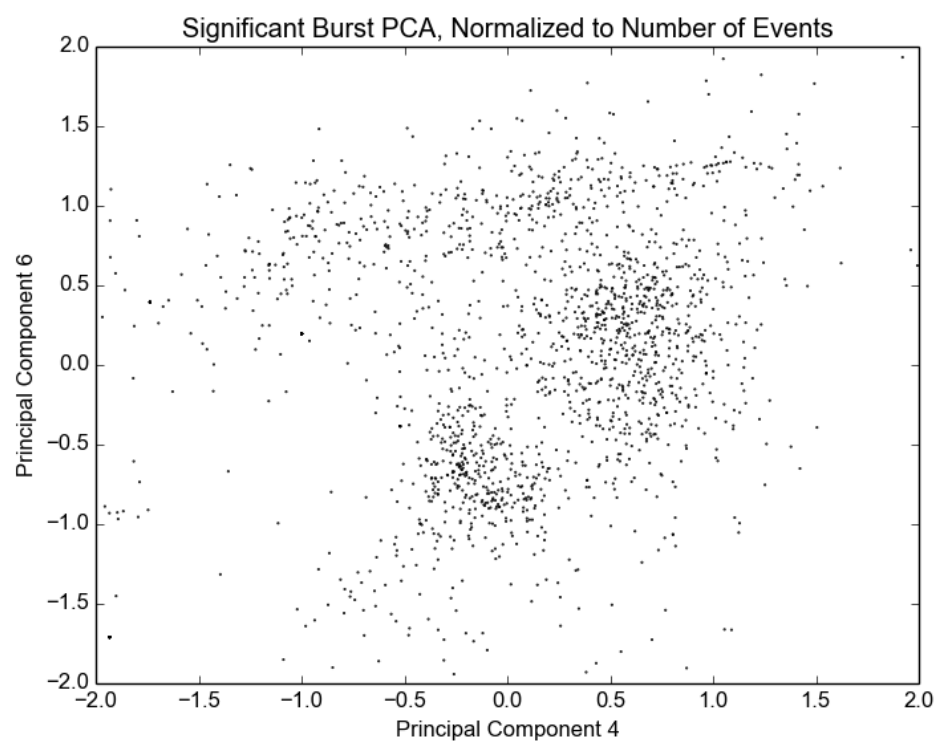
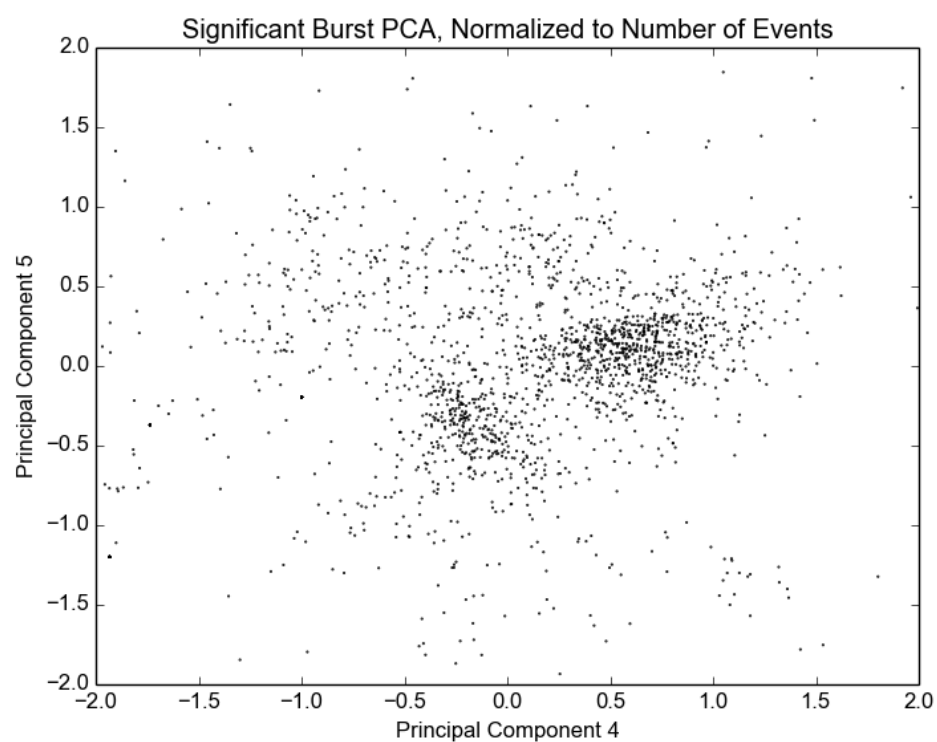


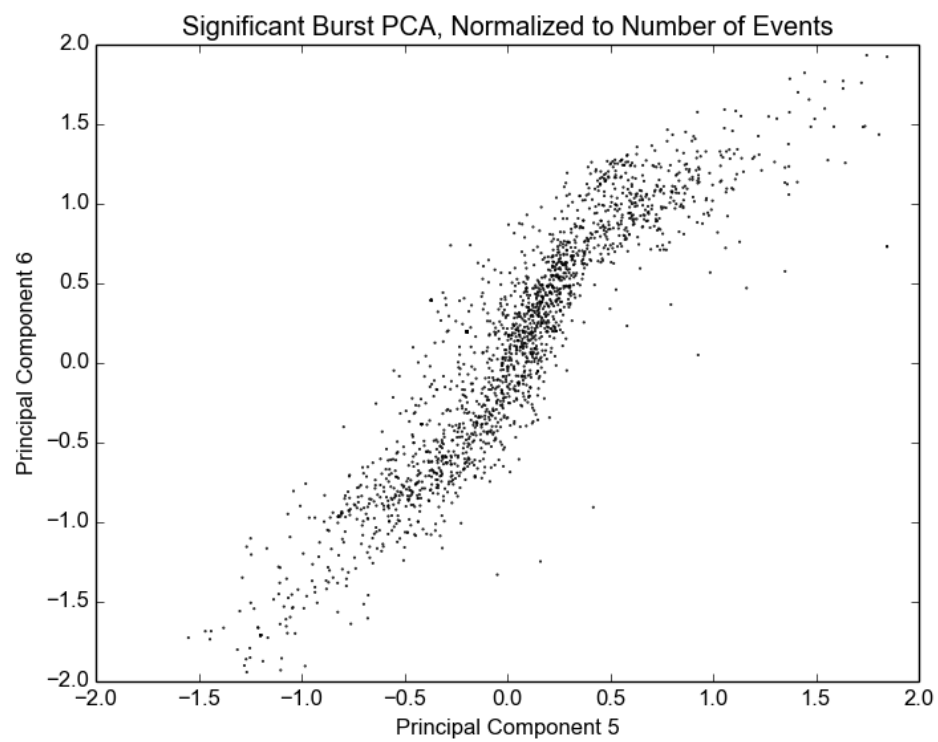
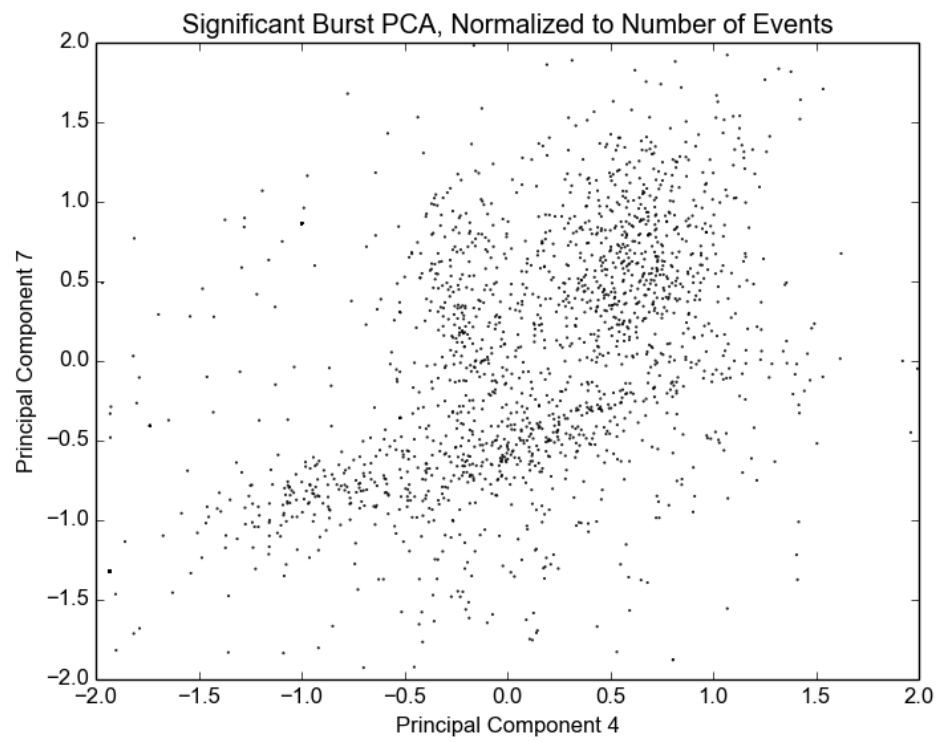


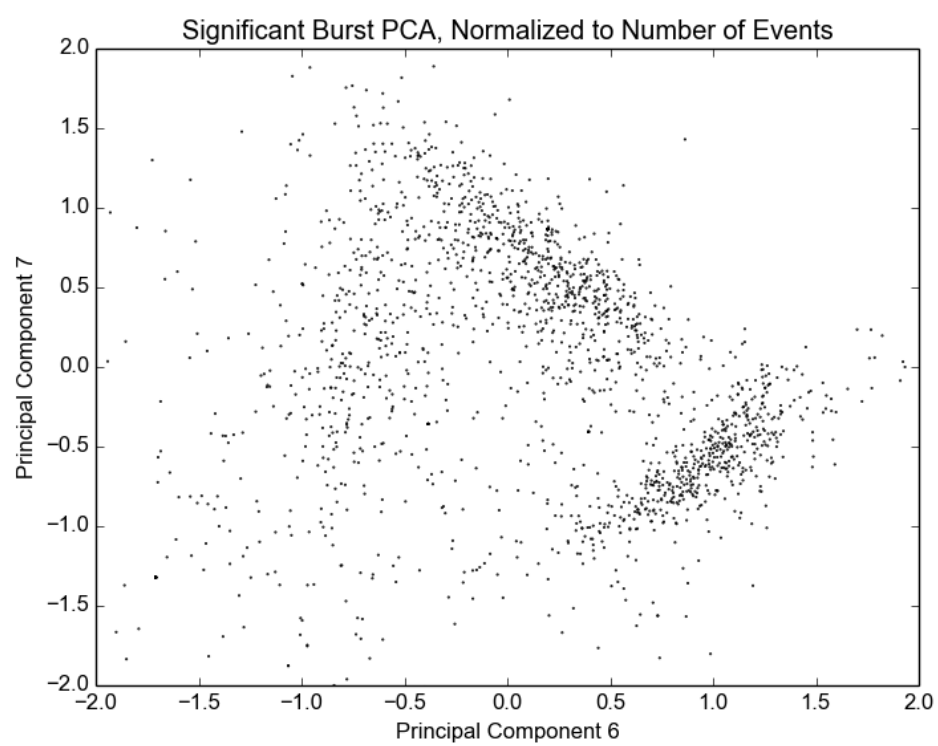
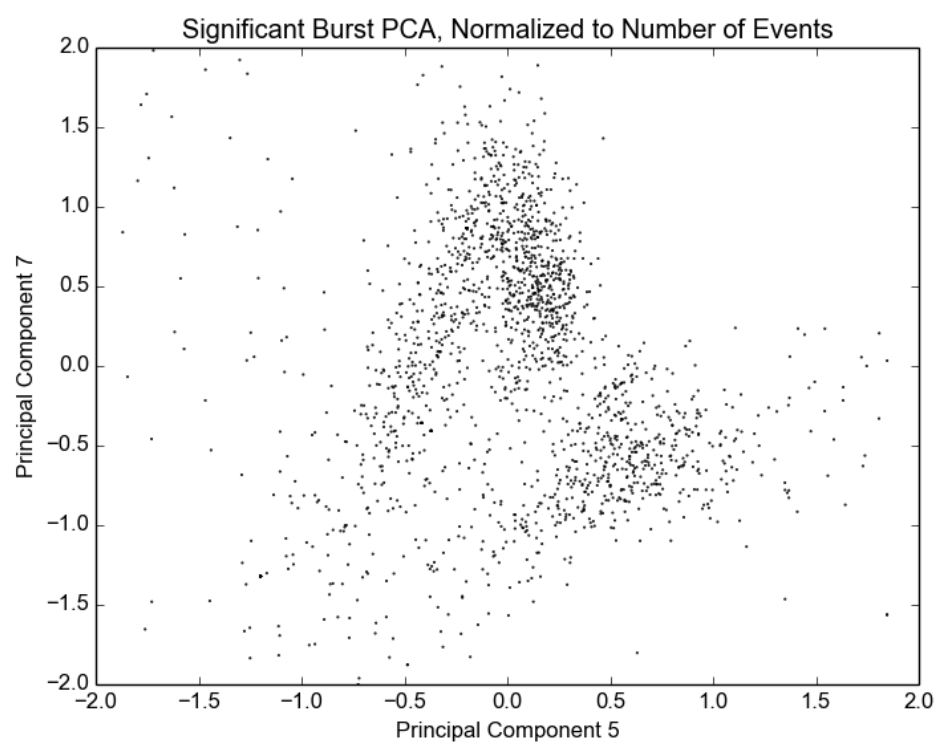




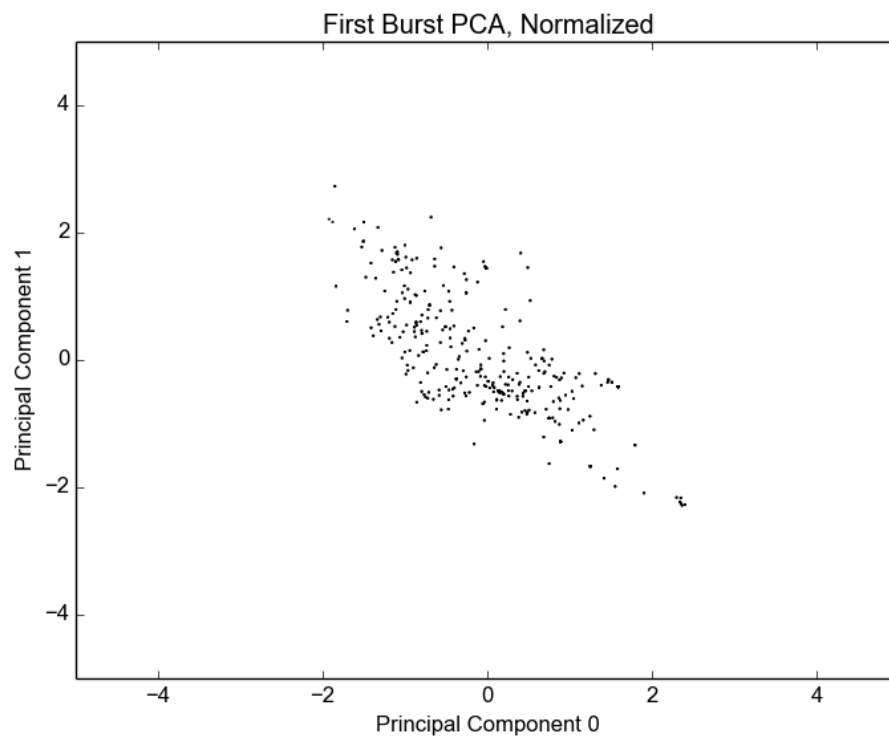


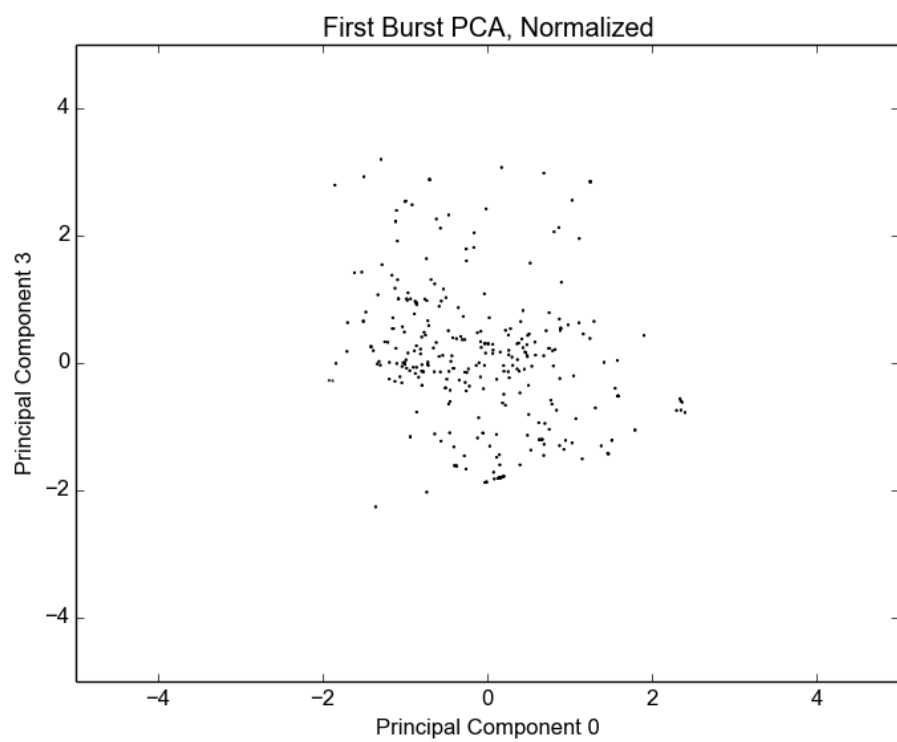
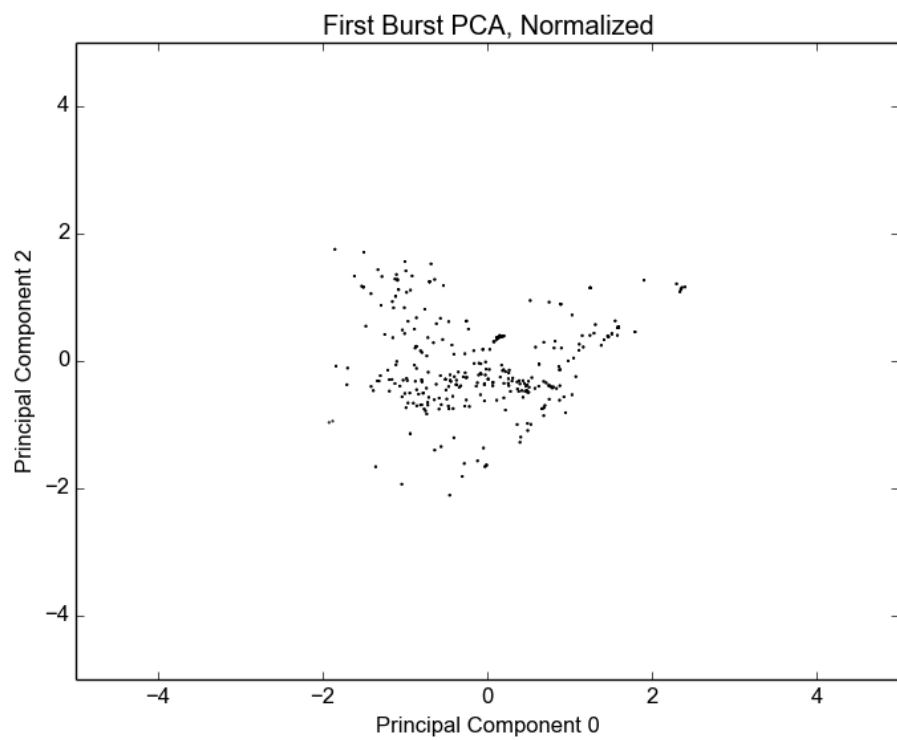


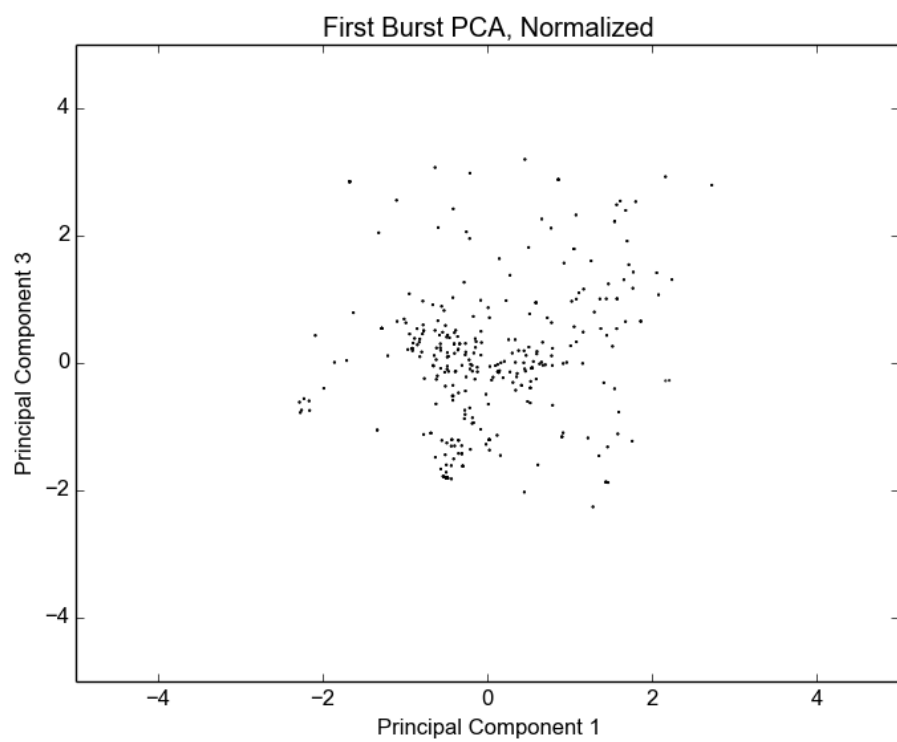
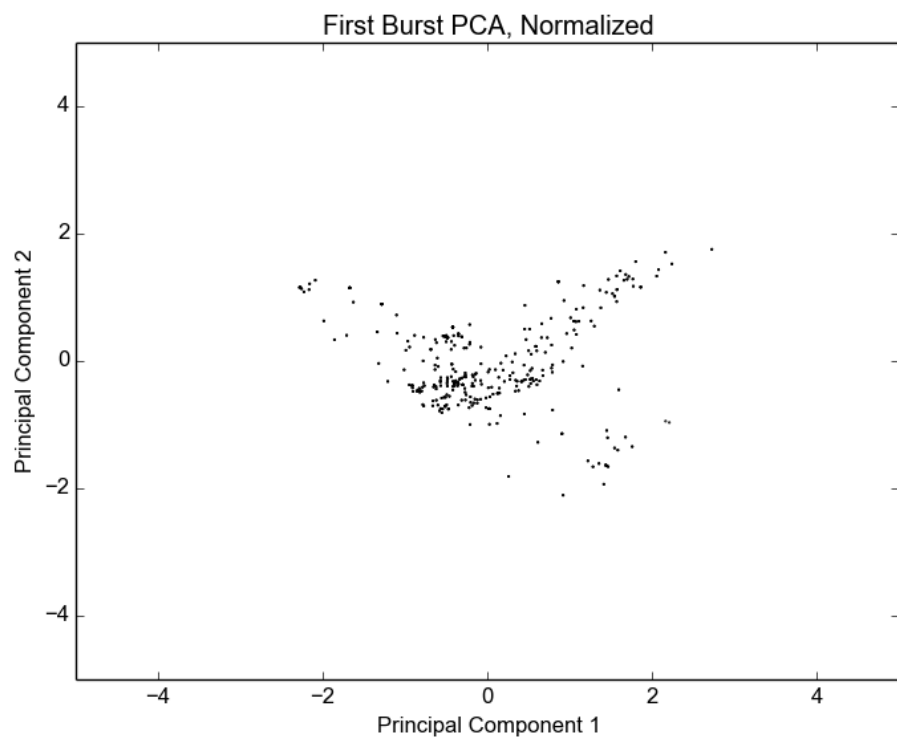


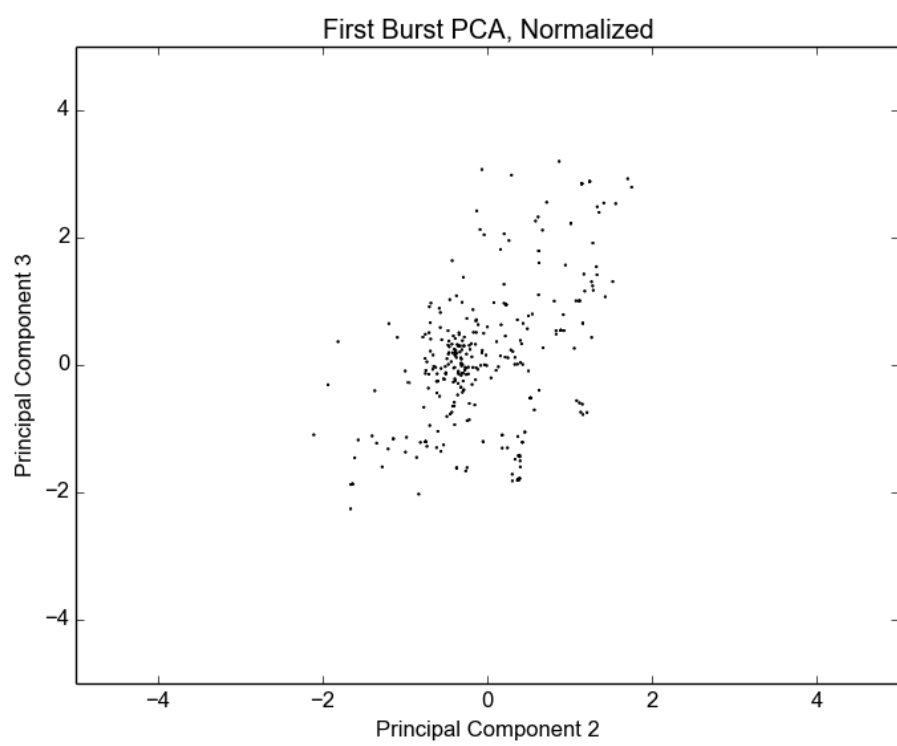


C.3 First Bursts PCA plots

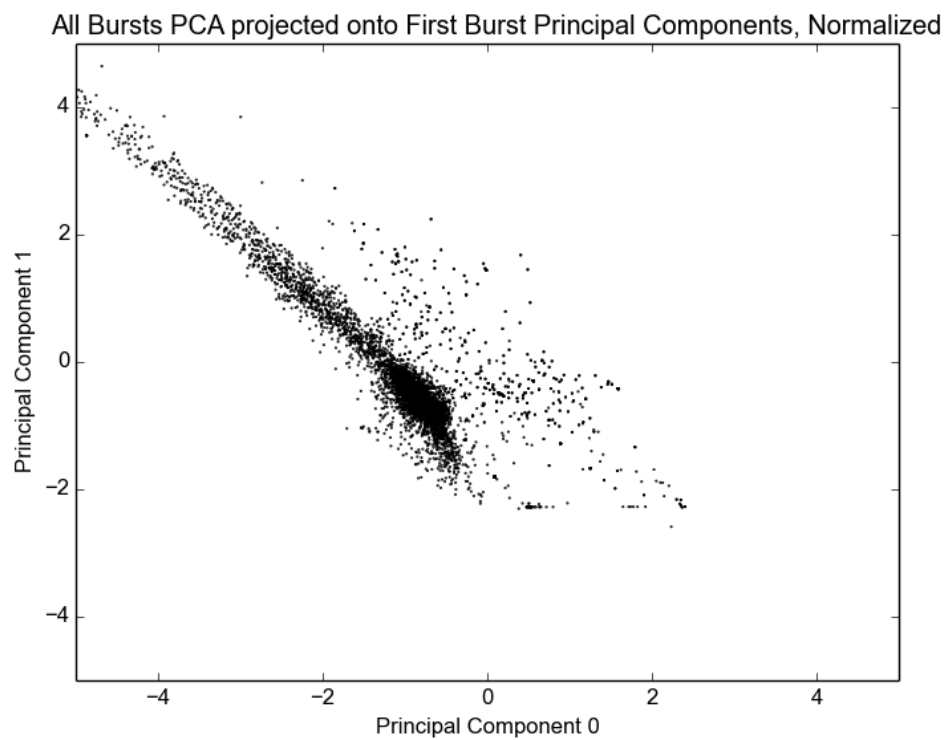




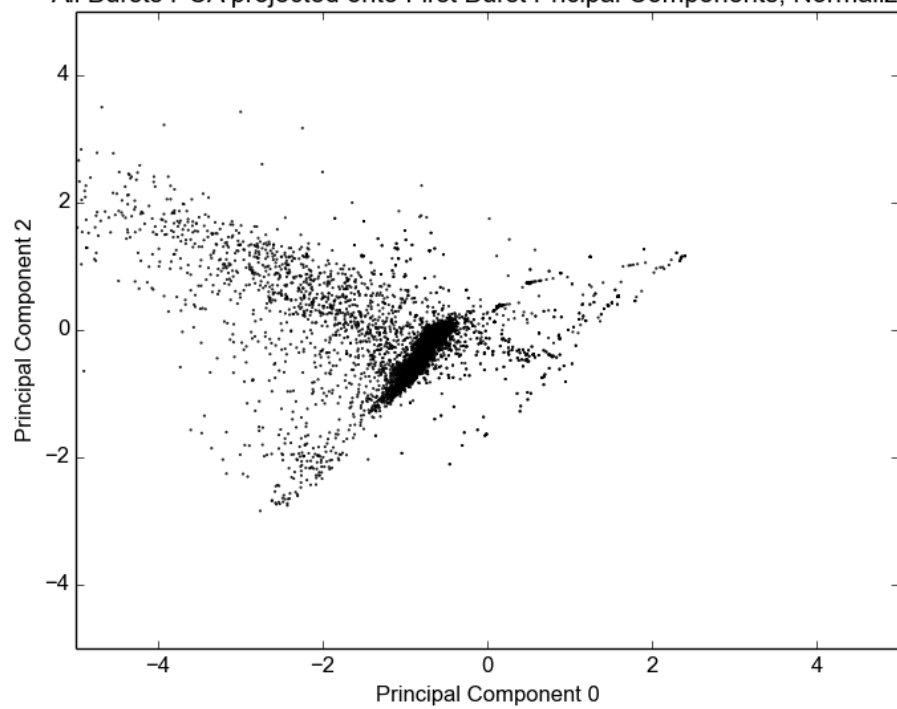




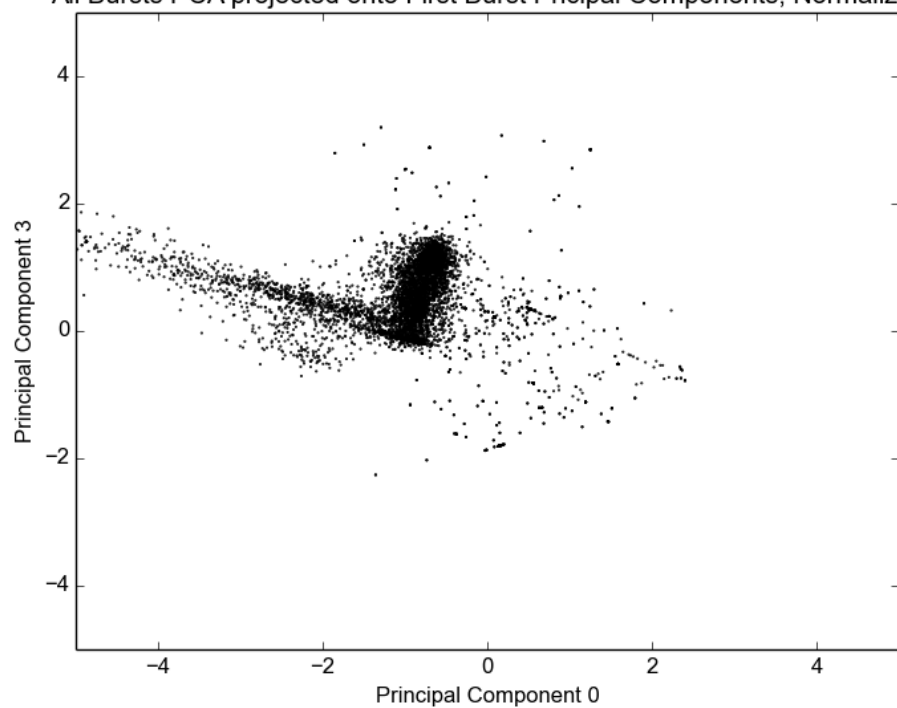
C.4 All Bursts Projected Onto First Bursts PCA plots



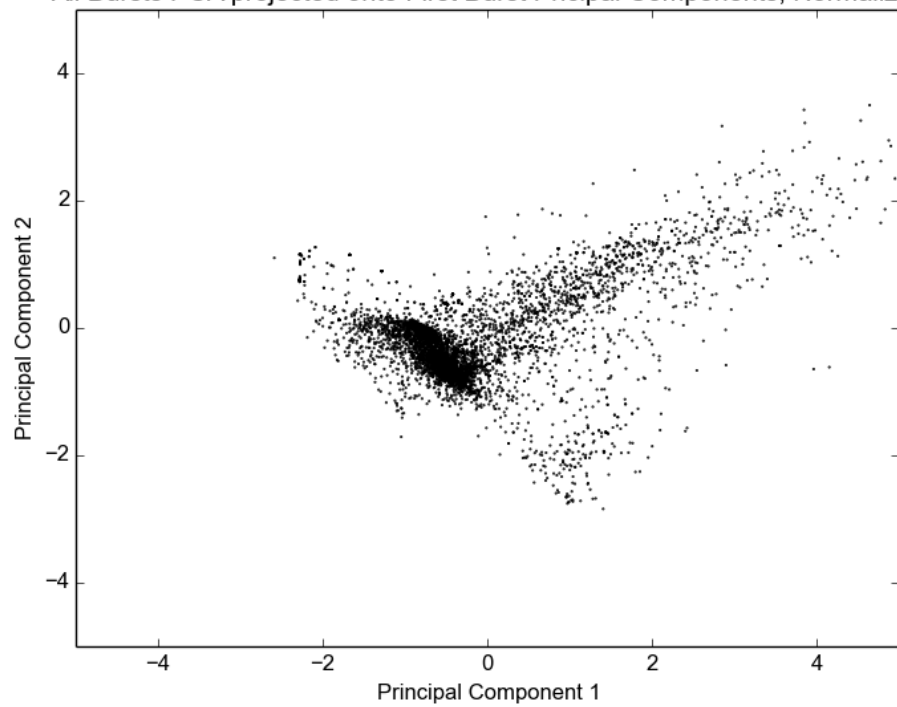
All Bursts PCA projected onto First Burst Pricpal Components, Normalized



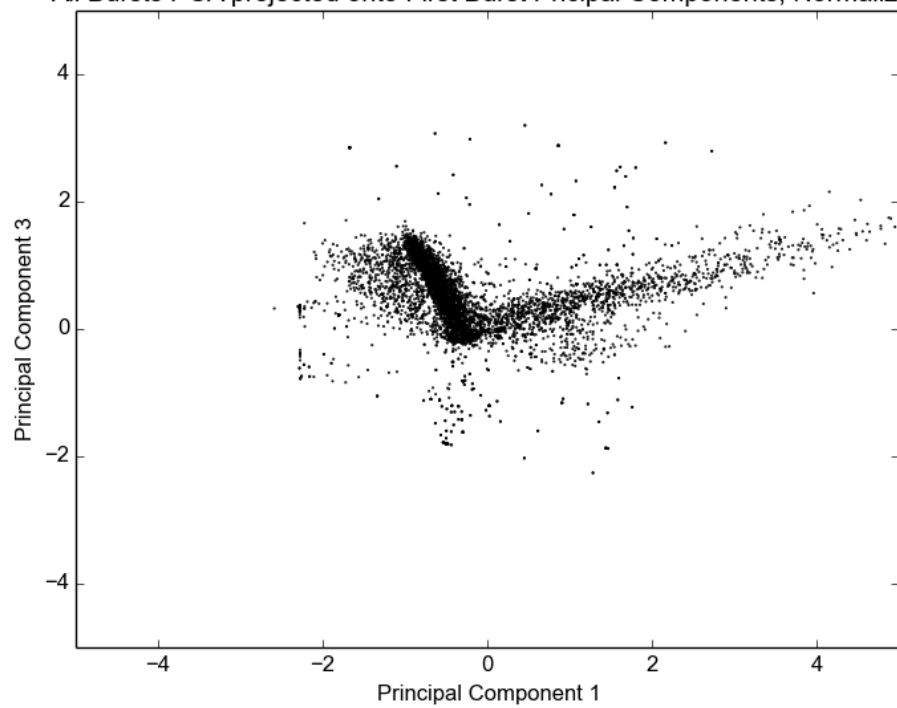
All Bursts PCA projected onto First Burst Pricpal Components, Normalized

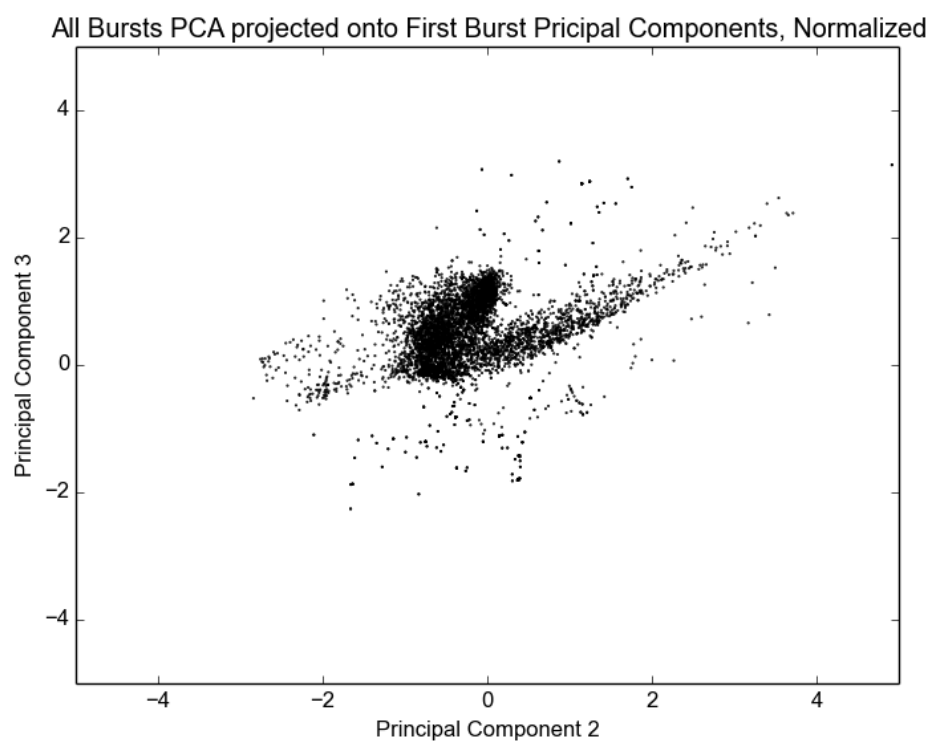


All Bursts PCA projected onto First Burst Pricpal Components, Normalized



All Bursts PCA projected onto First Burst Pricpal Components, Normalized





D Normalized Data Cleaning Flag Distribution Histograms

The total number of times a flag had been called per event in a burst file was plotted for every tenth burst file generated between May 2017 and February 2018 (Figures 45-53). The distribution of the number of events in a burst file from the same sample was also plotted in Figure 55.

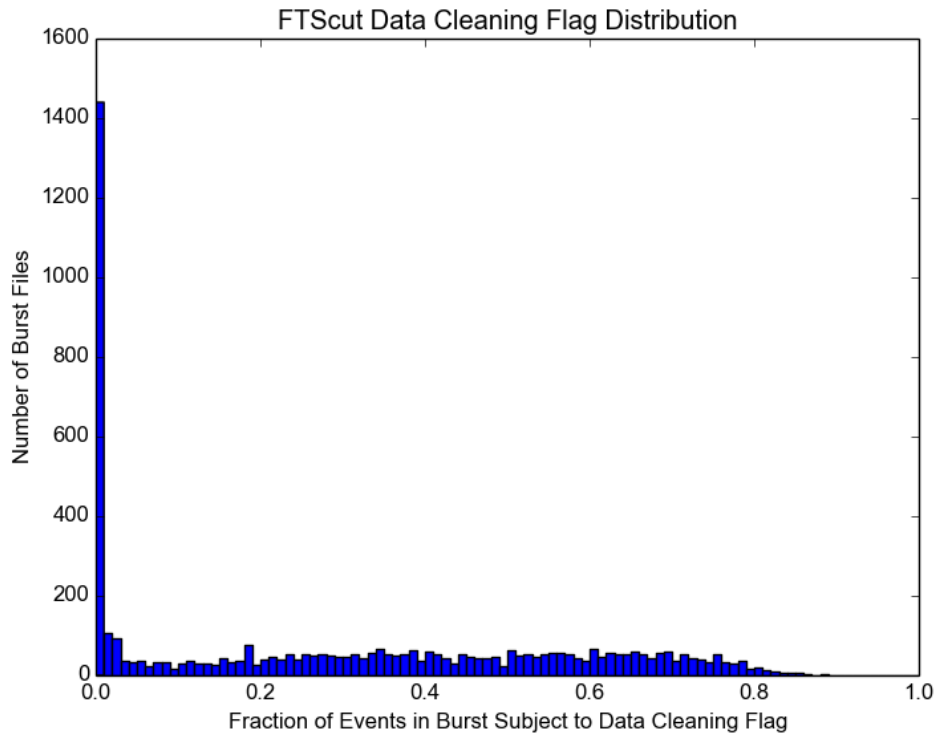


Figure 83: FTScut flags/event distribution - data distribution is \sim uniform with a peak at the Zeroth bin. Distribution tapers to zero above 0.75 flags/event.

By using this method the data sorts itself into a defined scale in which bin numbers represents the fraction of events in the burst file. The counts of in a specific bin refer to the number of burst files for which that specific ratio was true.

In this form, all of the distributions once again present with \sim singular peaks and less

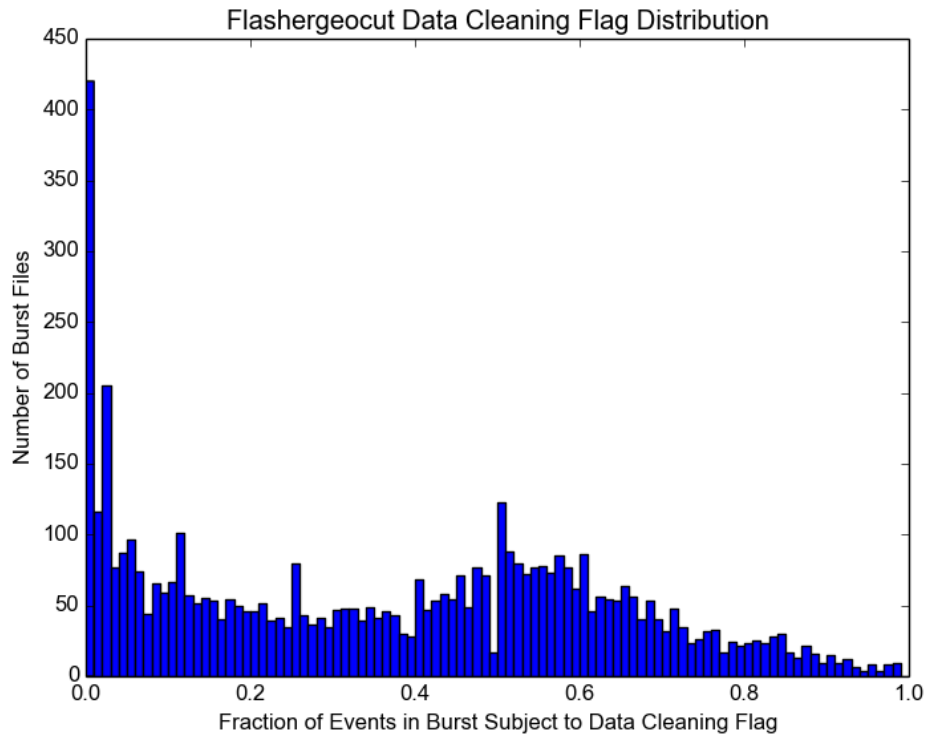


Figure 84: Flashergeocut flags/event distribution - data shows a normal peak about 0.5 flags/event with the zeroth bin as the most populated.

prominent tails. With the exception of the Qcluster distribution these peaks generally form at one end of the scale or the other. This indicates that a specific flag would generally be applied or not to \sim all of the events in the burst file.

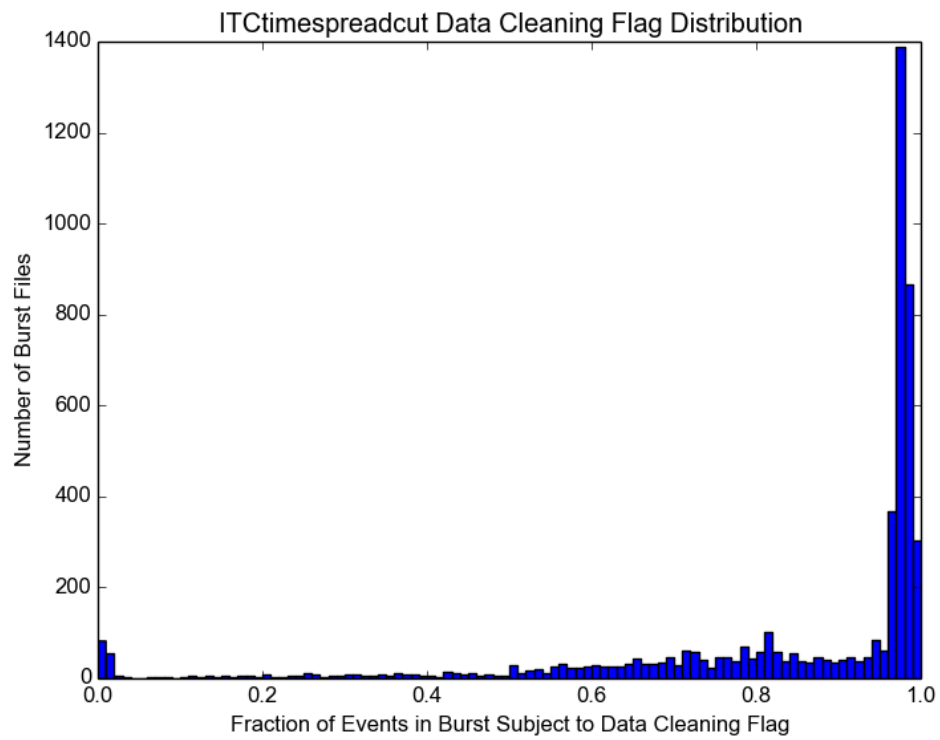


Figure 85: ITCtimespreadcut flags/event distribution - Peak at 0.97 flags/event with weak tail toward zero. Zeroth bin more populated than intermediate bins.

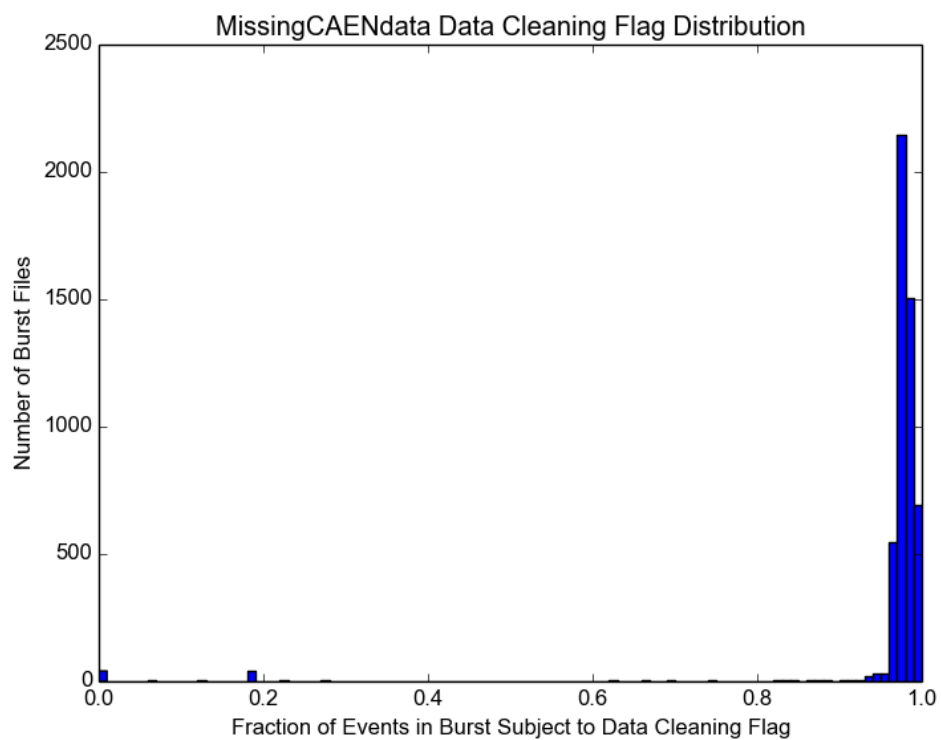


Figure 86: Missingcaendata flags/event distribution - Peak at 0.97 flags/event with weak tail toward zero..

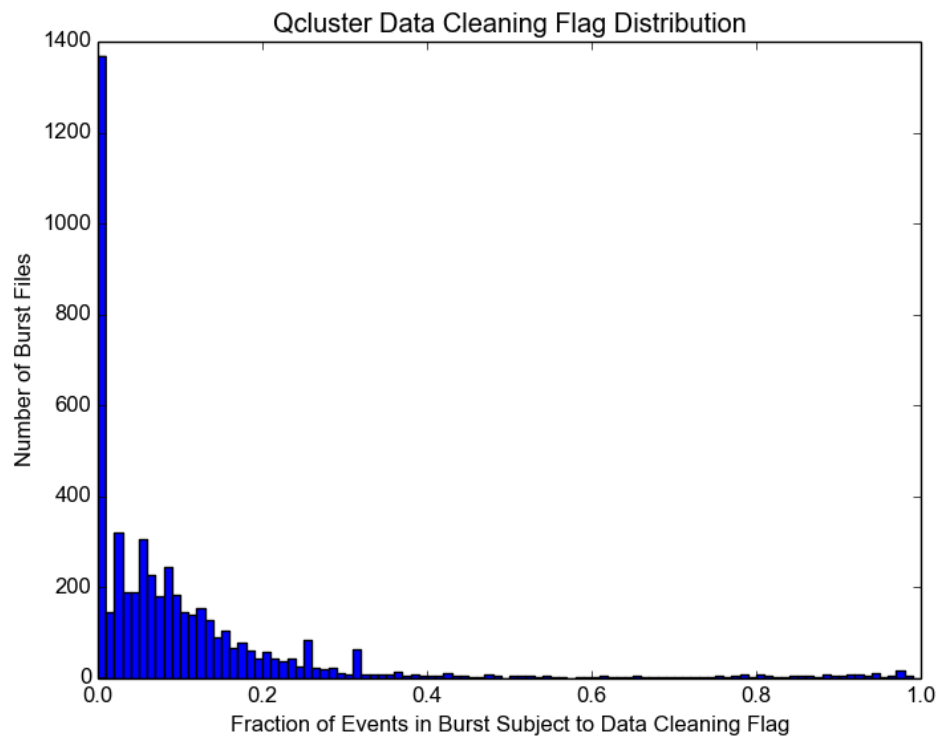


Figure 87: Qcluster flags/event distribution - Peak at zeroth bin. A "bump" can be seen centered about 0.08 flags/event.

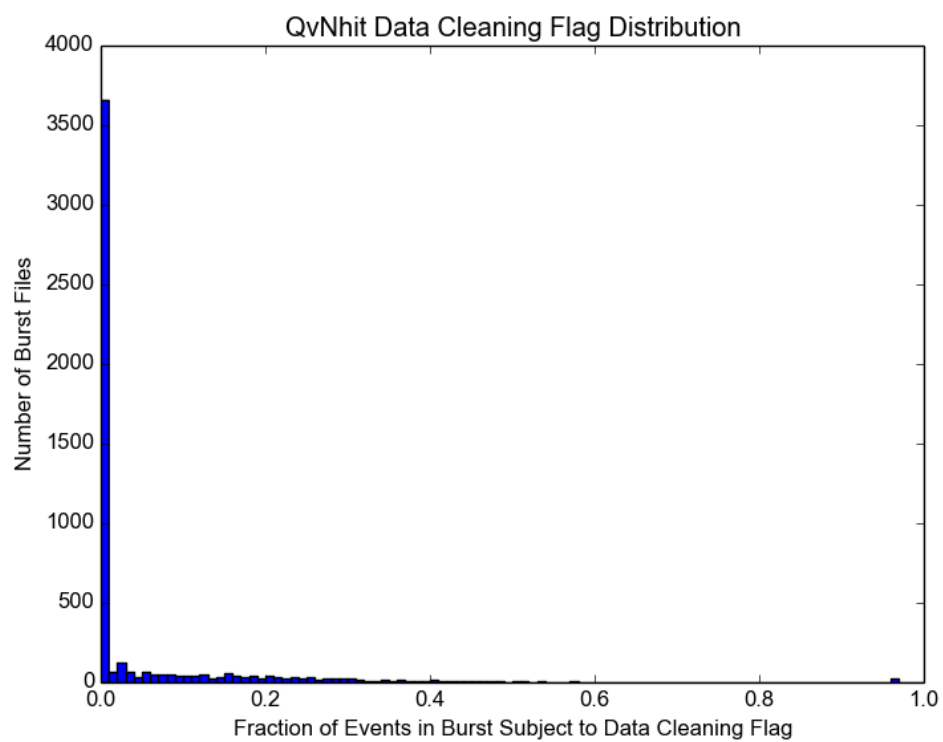


Figure 88: QvNhit flags/event distribution - Peak at zeroth bin with a weak tail toward the right. No discernible structure to the rest of the data distribution

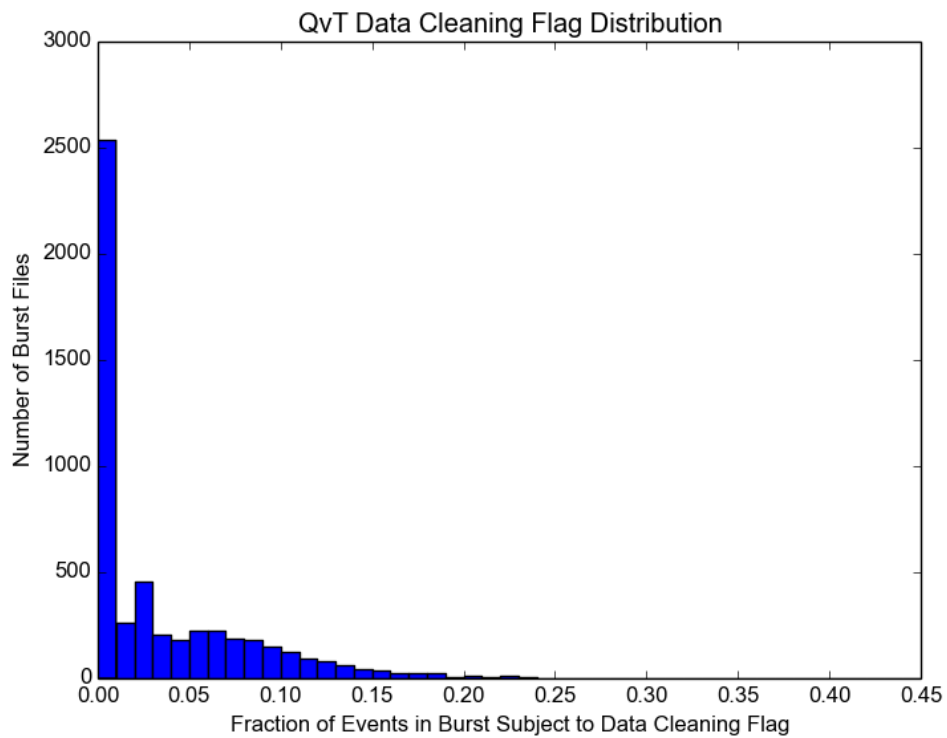


Figure 89: QvT flags/event distribution - Peak at zeroth bin. A "bump" can be seen centered about 0.06 flags/event. X scale reduced as bins above 0.25 flags/event are not populated.

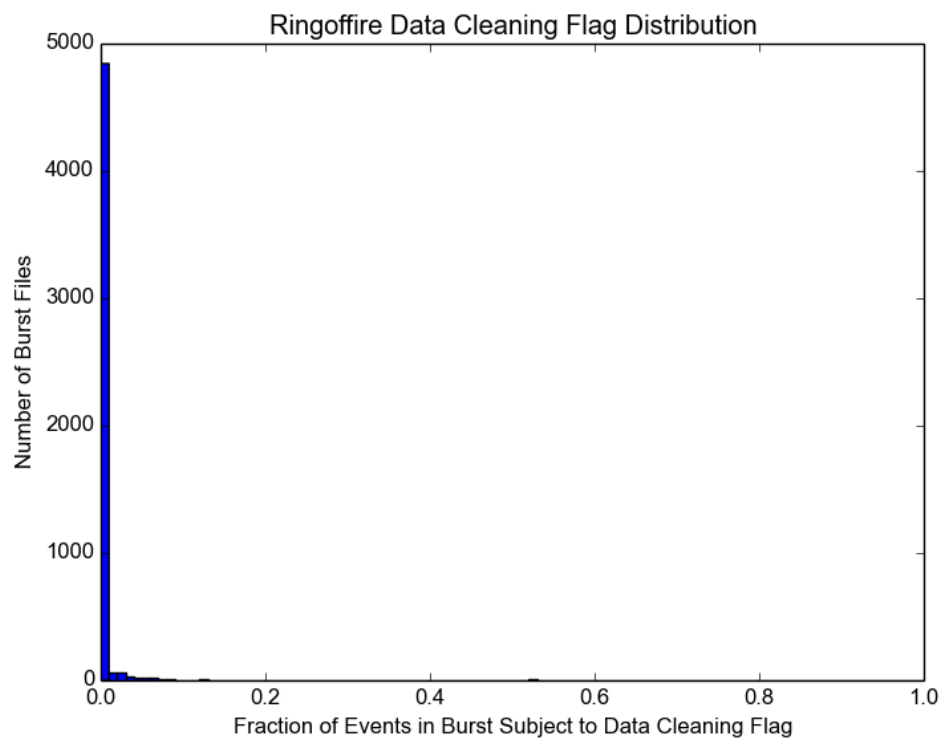


Figure 90: Ringoffire flags/event distribution - Peak at zeroth bin with a weak tail toward the right. No discernible structure to the rest of the data distribution

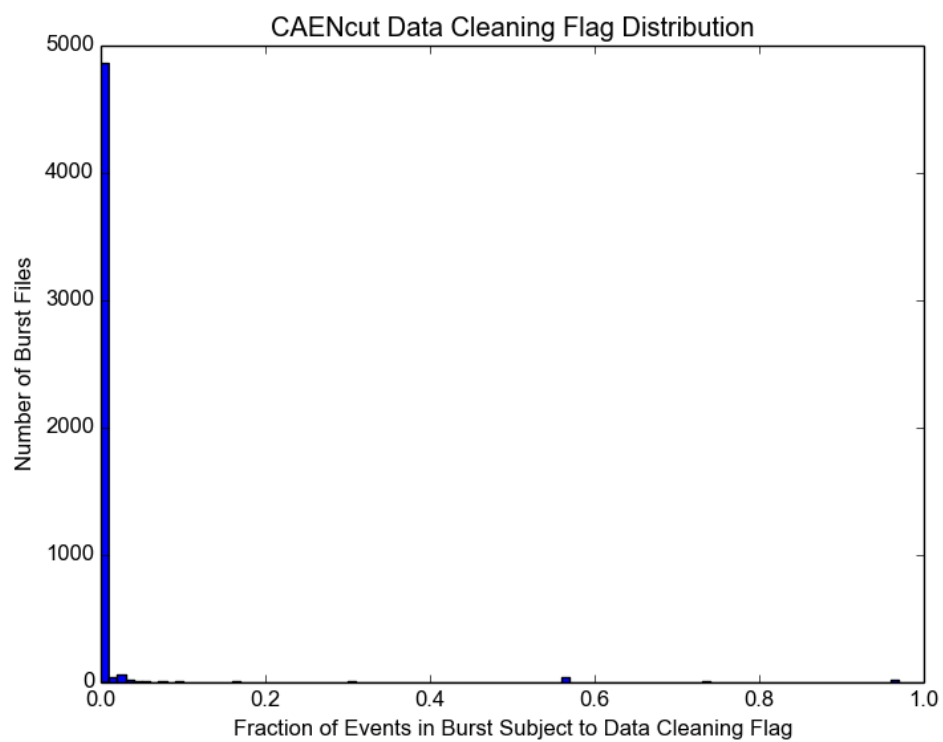


Figure 91: CAENcut flags/event distribution - Peak at zeroth bin with a weak tail toward the right. No discernible structure to the rest of the data distribution

E Data Analysis Scripts

E.1 Batch-wise Data Processing

This is the main bash script which is used to process a batch of burst files. It iterates over a list of burst files creating the rat macro tailored to run each burst file in the list. I call the ROOT macro to read out the data cleaning flags to file which are then processed into an individual vector by the python script. The vector is appended to a list and the number of events in the files are stored in a separate file to be used later.

```
#ls *.zdab > filenames.txt

cat filenames.txt | while read LINE;

do echo "$line"

scp /media/philip/Exthdd/Burst/${LINE} /home/philip/Desktop/work

firstline="/rat/physics_list/OmitAll true"

firstpasslastline="/rat/db/set DETECTOR geo_file \"geo/snoplus_water.geo\"

/run/initialize

/rat/proc datacleaning

/rat/procset add \"tpmuonfollowercut\"

/rat/procset pass 1

/rat/inzdab/read

exit"

echo "$firstline" > firstpass.mac
```



```

echo "/rat/inzdab/load ${LINE}" >> firstpass.mac

echo "$firstpasslastline" >> firstpass.mac


secondpassecondline="/rat/db/set DETECTOR geo_file

    \"geo/snoplus_water.geo\"

    /run/initialize

    /rat/proc calibratePMT

    /rat/proc datacleaning

    /rat/procset mask \"default_apply\"

    /rat/procset add \"tpmuonfollowercut\"

    /rat/procset pass 2

    /rat/proclast outroot"

secondpasslastline="/rat/inzdab/read

    exit"

echo "$firstline" > secondpass.mac

echo "/rat/inzdab/load ${LINE}" >> secondpass.mac

echo "$secondpassecondline" >> secondpass.mac

echo "/rat/procset file \"output.root\" \"\" >> secondpass.mac

echo "$secondpasslastline" >> secondpass.mac


source ~/snoing/install/env_rat-dev.sh

rat firstpass.mac

rat secondpass.mac

```

```
./ReadDC1 output.root

python vectors.py

echo "${LINE}" >> vectors.txt

rm ${LINE}

rm rat.*

rm tpmuonfollowercut*

done


#rm filenames.txt

#rm firstpass.mac

#rm output.root

rm rat.*

rm secondpass.mac
```

E.2 ROOT Data Cleaning Flag Extraction Macro

This macro is run in ROOT and extracts the number of events and data cleaning flag information from the processed burst file. These are written to separate files. A check is run after the fact to ensure that the number of events in a specific burst is matched to the specific data cleaning flag vector which is compiled by the python script.

```
////////////////////////////////////
/// \file ReadDCFlags_NoPlots.cc
///
/// \Program to read out data quality flags from a processing ROOT file
////////////////////////////////////

#include <RAT/DU/DSReader.hh>

#include <RAT/DU/Utility.hh>

#include <RAT/DU/DataCleaningBits.hh>

#include <RAT/DS/Entry.hh>

#include <RAT/DS/EV.hh>

#include <RAT/DS/DataQCFlags.hh>

#include <RAT/DS/BitMask.hh>

#include <RAT/DS/Meta.hh>

#include <TH1D.h>

#include <TCanvas.h>
```

```

#include <TStyle.h>

#include <TLegend.h>


#include <typeinfo>

#include <string>

#include <map>

#include <iterator>


#include <fstream>

#include <sstream>

#include <iostream>

#include <stdio.h>


using namespace std;

std::string writefile = "";


vector<string> known_cuts()
{

    // List of some known cuts that are not the TPMuonFollower.

    // Want to weed out the values so we're not printing so much.

    const char *vinit[] = {"prescalecut", "crateisotropy","qcluster",
                           "ringoffire","waterblindlow0","waterblindhigh0",

```

```

        "itctimespreadcut", "qvnhit", "flashergeocut", "owlcut",
        "zerozerocut", "qvt", "ftscut", "junkcut",
        "waterblindhigh4", "waterblindlow9"};

    int length = sizeof(vinit)/sizeof(vinit[0]);

    vector<string> vec(vinit, vinit+length);

    return vec;
}

```

```

bool isKnownDCFlag(string input)
{
    /* Compares a fed in DCFlag against the known_cuts vector. If in the
     * list, returns true.
     */

    bool isknown = 0;

    vector<string> knowncuts = known_cuts();

    for (int i = 0; i < knowncuts.size(); i++){
        if (input == knowncuts[i])
            isknown = 1;
    }

    return isknown;
}

```

```

int main(int argc, char** argv)

```

```

{

    const string& filename = string(argv[1]);

    // Read root files

    RAT::DU::DSReader dsReader( filename );

    cout << dsReader.GetEntryCount() << endl;

        Double_t v1;

        std::ofstream outlog;

        outlog.open("enums.txt", std::ios_base::app);

        outlog << s5 << endl;

        outlog << dsReader.GetEntryCount() << endl;

    cout << "CHECKING FOR TPMFC IN FILE " << filename << endl;

    RAT::DU::DataCleaningBits rDataCleaningBits =

        RAT::DU::Utility::Get()->GetDataCleaningBits(); // To get the data

        cleaning bits

    const char *s4 = writefile.c_str(); //write to .txt

    // Loop through entries in rootfile

    for( size_t iEntry = 0; iEntry < dsReader.GetEntryCount(); iEntry++ ){

        if( iEntry % 10000 == 0 ){ cout << "iEntry: " << iEntry << " / " <<

            dsReader.GetEntryCount() << endl; }

        const RAT::DS::Entry& rDS = dsReader.GetEntry( iEntry );

```

```

// Look at each triggered event in the entry

for( size_t iEV = 0; iEV < rDS.GetEVCount(); iEV++ ){

    if( iEV != 0 )

        continue;

    const RAT::DS::EV& rEV = rDS.GetEV( iEV );

    const RAT::DS::DataQCFlags& rDataQCFlags = rEV.GetDataCleaningFlags();

    const RAT::DS::Meta& rMeta = dsReader.GetMeta();

    const UInt_t pass = rMeta.GetCurrentPass();


    // Get Nhits for all and OWL tubes

    int calnormnhits = rEV.GetCalPMTs().GetAllCount();

    int owlhits = rEV.GetCalPMTs().GetOWLCount();

    int neckhits = rEV.GetCalPMTs().GetNeckCount();


    // Check flags exist

    if(!(rDataQCFlags.ExistFlags(pass))) // Checks flags exist

        continue;

    // Flags exist! For this event, get the flags

    const RAT::DS::BitMask& rBitMaskFlags = rDataQCFlags.GetFlags(pass);

    const RAT::DS::BitMask& rBitMaskApplied = rDataQCFlags.GetApplied(pass);

```

```

// Loop through bits

for( map<size_t, string>::iterator iDCBits =

    rDataCleaningBits.GetInverseMapBegin();

    iDCBits != rDataCleaningBits.GetInverseMapEnd(); iDCBits++ ){

    Bool_t rApplied = rBitMaskApplied.Get( iDCBits->first );

    if( !rApplied ) // Checks cut has been applied

        continue;

    Bool_t rFlag = rBitMaskFlags.Get( iDCBits->first );

    if( rFlag )

        continue;

    string cutName = rDataCleaningBits.GetBitName( iDCBits->first

        ).c_str();

    if (calnormnhits > 100) {

        Double_t v1;

        std::ofstream outlog;

        outlog.open("filelog.txt", std::ios_base::app);

        outlog << s4 << endl;

        outlog << iEntry << " , " << cutName << endl;

        cout << "# total inward pmts hit:" << calnormnhits << endl;

        cout << "# total owl pmts hit:" << owlhits << endl;

        cout << "# total neck pmts hit:" << neckhits << endl;

```



```
    } // bit loop

    // Divide cut events by denominator to get sacrifice
} // EV

} // Entry
}
```

E.3 Data Cleaning Flag Vector Assembly

This python script iterates over the text file generated by the ROOT macro and assembles it's entirety into numerical values associated with the specific burst. The goal was to retain as many data cleaning flags as possible. This list contains all flags which were read out from the ROOT macro as well as a few other which were not populated

```
import os

entryNum = ( ' ' )

vector = []

with open("filelog.txt","r") as file:

    for line in file:

        if line.strip():

            fields = line.split(' , ')

            check = fields[1]

            entryid = int(fields[0])

            if entryid < entryNum:

                entryNum = entryid

                vector=[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]

                if fields[1]==('caencut\n'):

                    vector[0]=vector[0]+1;

                elif fields[1]==('ftscut\n'):
```

```

        vector[1]=vector[1]+1;

elif fields[1]==('flashergeocut\n'):

        vector[2]=vector[2]+1;

elif fields[1]==('itctimespreadcut\n'):

        vector[3]=vector[3]+1;

elif fields[1]==('neckcut\n'):

        vector[4]=vector[4]+1;

elif fields[1]==('qcluster\n'):

        vector[5]=vector[5]+1;

elif fields[1]==('qvnhit\n'):

        vector[6]=vector[6]+1;

elif fields[1]==('qvt\n'):

        vector[7]=vector[7]+1;

elif fields[1]==('ringoffire\n'):

        vector[8]=vector[8]+1;

elif fields[1]==('caencut\n'):

        vector[9]=vector[9]+1;

elif fields[1]==('crateisotrop\n'):

        vector[10]=vector[10]+1;

elif fields[1]==('prescalecut\n'):

        vector[11]=vector[11]+1;

elif fields[1]==('zerozerocut\n'):

        vector[12]=vector[12]+1;

```

```

elif fields[1]==('junkcut\n'):

    vector[13]=vector[13]+1;

elif fields[1]==('muontag\n'):

    vector[14]=vector[14]+1;

elif fields[1]==('owlcut\n'):

    vector[15]=vector[15]+1;

elif fields[1]==('tpmuonfollowercut-short\n'):

    vector[16]=vector[16]+1;

elif fields[1]==('tpmuonfollowercut-long\n'):

    vector[17]=vector[17]+1;

elif fields[1]==('nothingcut\n'):

    vector[18]=vector[18]+1;

elif fields[1]==('nhitcut\n'):

    vector[19]=vector[19]+1;

else: print(fields[1])

log = open("vectors.txt","a")

print >> log, vector

os.remove("filelog.txt")

```

E.4 Vector Normalization Script

This python script iterates over the list of vectors and the list of total events in bursts matching them together and outputting a vector which consists of vectors of the form flags/event.

```
with open("vectors.txt","r") as file1:

    line_num = 0

    for line in file1:

        line_num += 1

        stripl = line.strip("(")

        nobra = stripl[:-2]

        split = nobra.split(", ")

        line2_num = 0

        with open("ENumbers.txt","r") as file2:

            for line in file2:

                events = float(line)

                line2_num +=1

                if line2_num == line_num:

                    a = round(float(split[0]) / events,5)

                    b = float(split[1]) / events

                    c = float(split[2]) / events

                    d = float(split[3]) / events

                    e = float(split[4]) / events

                    f = float(split[5]) / events
```

```
g = float(split[6]) / events
h = float(split[7]) / events
i = float(split[8]) / events
j = float(split[9]) / events
k = float(split[10]) / events
l = float(split[11]) / events
m = float(split[12]) / events
n = float(split[13]) / events
o = float(split[14]) / events
p = float(split[15]) / events
q = float(split[16]) / events
r = float(split[17]) / events
s = float(split[18]) / events
t = float(split[19]) / events

log = open("result.txt", "a")

list = [a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r,
        s, t]

print >> log, list
```

E.5 Python Script to Create Desired Vector Components

This script reduces the dimensions of the vectors used for PCA to only those components which we are interested in:

```
import numpy as np

with open("result.txt","r") as file:

    for line in file:

        vector = [0,0,0,0,0,0,0,0,0,0]

        stripl = line.strip("(")

        nobra = stripl[:-2]

        split = nobra.split(", ")

        vector =

            [float(split[1]),float(split[2]),float(split[3]),float(split[0]),

float(split[5]),float(split[6]),float(split[7]),float(split[8]),float(split[9])]

        print vector # NOTE THE FIRST VALUE IN ARRAY IS B=0!!!!!! IF YOU LOOK

            FOR ELEMENT 1 IT WILL BE THE SECOND VALUE

        log = open("element.txt","a")

        print >> log, vector
```

E.6 Principal Component Analysis Script

This script is based off of the example for the ROOT LINTRA package which has functionality built in for principal component analysis. It reads in the normalized data cleaning flag vectors and outputs a ROOT script generated to transform the data onto the principal component axes. At this point the data can be plotted after being transformed using your choice of software.

```
#include "TPrincipal.h"

#include <iostream>

#include <string>

#include <sstream>

#include <fstream>

void principal()
{

    TPrincipal* principal = new TPrincipal(9,"ND"); //CHANGE TO NUMBER OF
        COMPONENTS

    //load the text file

    std::ifstream in("element.txt");
```



```

if(!in){ // check for valid file

    std::cout<< "Error" << std::endl;

}

std::string line;

while(getline(in, line)){

    // Following items are to keep count, etc, do not change

    size_t pos1 = 1;

    size_t pos2;

    double finalNum;

    std::string value;

    //double array to hold numbers from each row

    Double_t* dbl= new Double_t[8]; // CHANGE THIS NUMBER TO ONE LESS THAN

        NUMBER OF COMPONENTS

    //Loop to pull out numbers from text

    for (int x=0; x<=7; x++){ // CHANGE X<= TWO LESS THAN NUMBER OF

        COMPONENTS

    pos2 = line.find(", ", pos1);

    value = line.substr(pos1, (pos2-pos1));

```

```

//converts string to double, put into array

std::stringstream number(value);

number >> finalNum;

dbl[x] = finalNum;


//increase pointer location to loop through row

pos1 = pos2+2;

}


//collect last number in line, format changed

pos2 = line.find("]", pos1);

pos2 = pos2;

value = line.substr(pos1, (pos2-pos1));


// Turn last string number into double

std::cout<< x << std::endl;

std::stringstream number(value);

number >> finalNum;

dbl[8] = finalNum; //CHANGE NUMBER TO ONE LESS THAN NUMBER OF COMPONENTS


//dbl is array containing a SINGLE LINE of your document. it CHANGES
every line.

```

```

// Add array to principal
principal->AddRow(dbl);
}

//delete array since in principal
delete [] dbl;

// Do the actual analysis
principal->MakePrincipals();

// Print out the result on
principal->Print();

// Test the PCA
principal->Test();

// Make some histograms of the orginal, principal, residue, etc data
principal->MakeHistograms();

// Make two functions to map between feature and pattern space
principal->MakeCode();

```

```
// Start a browser, so that we may browse the histograms generated  
  
// above  
  
TBrowser* b = new TBrowser("principalBrowser", principal);  
  
}
```
